CHILDREN'S UNDERSTANDING OF QUANTITY AND THEIR ABILITY TO USE GRAPHICAL INFORMATION

Carmen Josefina Cividanes-Lago
Somerville College

A thesis submitted for the degree of Doctor of Philosophy

University of Oxford, Trinity Term, 1993

CHILDREN'S UNDERSTANDING OF QUANTITY AND THEIR ABILITY TO USE GRAPHICAL INFORMATION

Carmen Josefina Cividanes-Lago
Somerville College

A thesis submitted for the degree of Doctor of Philosophy
University of Oxford, Trinity Term, 1993

Abstract

This investigation concerns the ways in which young children (ages 5 to 8) compare quantities and how they work out the difference between them.

The experiments involved children's understanding of mathematical problems and their ability to make use of graphical information in such problems. Each child was shown a series of illustrations, each representing two sets of quantities where the numerical difference was represented either discontinuously or continuously. The children were asked Equalize and Compare questions about each illustration and had to choose the correct answer from the set which represented the choice stimuli. Children's use of strategies was observed.

In Experiment 1 (5-to-8-year-olds), only the younger children (5-to-6-year-olds) were observed to perform much more accurately on the Equalize-type question than on the Compare in both discontinuous and continuous conditions. The 7-to-8-year-olds reached a ceiling effect in performance, suggesting that by this age they can already deal with different types of arithmetic problems and with different types of graphical information.

Experiment 2 (5-to-6-year-olds) repeated the first experiment presenting the graphical information on a microcomputer, but the discontinuous and continuous conditions were subdivided on the basis of the use of the comparative term "more" or "less". Children are helped significantly by the use of discontinuous material and by the use of "more" in Equalize-type questions only. These results did not support those of Experiment 1 where the Equalize and Compare difference was significant with both discontinuous and continuous material.

Experiment 3 introduced part-whole manipulations in order to find out why Compare questions are more difficult to solve than Equalize questions. Five-to-6-year-olds' performance on Compare word problems was not affected by this type of manipulation.

Experiment 4 explored the Equalize and Compare difference by presenting the material in a story-telling context. Again, the 5-to-6-year-olds' performance on Compare word problems was not affected by this type of manipulation. However, Equalize questions were helped by the use of the comparative term "more", as in Experiments 2 and 3, and by the presentation of discontinuous material, as in Experiment 2.

Experiments 5 and 6 explored children's (5-to-8-year-olds) performance on Equalize- and Compare-type questions using spatial imagery manipulations. Experiment 5 involved manipulations of display in order to examine children's relative ease with Equalize word problems. Again, children's performance was not affected by this type of manipulation. In addition to the display manipulations, Experiment 6 introduced different level manipulations. However, in this experiment, the comparative pair was not represented in the choice stimuli. Children's performance on Compare word problems improved. There was no sign of the Equalize and Compare distinction which may be due to the fact that there was no representation of the comparative pair.

The results show that the Equalize and Compare difference is due to a combination of their inherent structural and linguistic factors. Furthermore, the difficulty children have with Compare word problems is non-number-specific, but their relative ease with Equalize word problems is number-specific. Such type results indicate that children represent these two problems very differently.

This thesis contains approximately 94,000 words.

These studies were supported with the help of grants from the Overseas Research Students' Scheme, administered by the Committee of Vice-Chancellors and Principals of the Universities of the United Kingdom and from the Janet Watson Bursary, administered by Somerville College, University of Oxford, England.

Parts of this thesis were presented in the following conferences:

The British Psychological Society's Annual Conference, 31 March - 3 April 1989, University of St. Andrews, Fife, Scotland;

International School Psychology Association's XIVth Colloquium, 24-28 July 1991, University of Minho, Braga, Portugal;

British Society for Research into Learning Mathematics, 4 April 1992, University of Oxford, Oxford, England.

To my grandmother,

Josefina D. Riera Bengoechea

To my father,

Adolfo Cividanes Freiría

To my mother,

Carmencita Lago Riera

To my sister,

Marie Angelique

and brother,

Adolfo

Acknowledgements

Peter E. Bryant	My supervisor, without whom it would have never been the same.
Daphne M. S. D. Park	Past Principal of Somerville College, for being supportive throughout these years.
Nancy C. Waugh	My college advisor, for her concern.
Headmasters Teachers Children at the following Oxford Schools: Botley C.E. Primary, North Hinksey C.E. Primary, St. Barnabas' C.E. First, and West Oxford County First.	For being cooperative and participating in the studies.
Peter L. N. Naish	For helping me with the computer programming on the Apple II.
F. H. C. Marriott	For his useful comments and statistical advice.
Erick L. Suárez Pérez	For helping me with the computer graphics and continued statistical advice.
David M. Kennedy	For his patience with computer queries.
Gilbert I. Jenkins	For his editorial advice.

Lynette L. Bradley
Prajnaparamita DasGupta
Hourieh Dajani
Morag E. N. MacLean
Renata F. I. Meuter
Rachel C. Norris
Mary Sissons-Joshi
Mai Ahmed Zaki Yamani

For their friendship.

Betty M. Hammond
and her colleagues
at the Department of
Experimental Psychology,
University of Oxford.

For their willingness to help.

TABLE OF CONTENTS

ABSTRACT

This investigation concerns the ways in which young children (ages 5 to 8) compare two unequal quantities and how they work out the difference between them in word problems. Whereas this may seem to involve a simple mathematical operation, this thesis proposes that the operation required is a complicated one. The findings are relevant to understanding how a child's mathematical abilities develop.

Word problems are stories that present a mathematical problem and involve the comparison of two unequal quantities. A child solves a word problem by working out the difference between the two unequal quantities embedded as relevant information within the word problem. A review of the literature (see Chapter 1) identifies two types of word problem which involve the same mathematical operations, and yet children seem to have difficulty with one and not so much with the other.

Compare problems engage the child in analyzing the composition of two distinct quantities to determine if there is a difference between these and, if so, what that difference is. These problems may involve a "static" relationship between quantities. Static relations embody conceptual knowledge about comparisons involving sets of objects and do not pose a transformation between the quantities. An example of a Compare word problem is the following:

Mary has 6 books. John has 4 books.
How many more books than John does Mary have?

In Equalize problems two quantities are compared to discover the amount by which one of the quantities must be changed to make it equal to the other. Thus, a quantity transformation may be achieved by involving an

"action" relationship where addition or subtraction make one quantity equal to the other. Action relations embody actions that cause increases or decreases in some quantity and pose a transformation within the quantities. An example of an Equalize word problem is the following:

Mary has 6 marbles. John has 4 marbles.
How many more marbles does John need,
to have as many marbles as Mary?

Most children (ages 5 to 8) are able to perform Equalize problems correctly, and yet fail to accomplish the same with Compare problems. The literature contains two hypotheses regarding this difference between Equalize and Compare word problems. The first (Riley, Greeno, and Heller,1983) deals with the structural aspect of these types of problem and makes a claim for the difference between static versus action relationships, as previously mentioned. The second hypothesis (Hudson,1983) deals with the linguistic difficulties of Compare problems and involves children's failure to understand the comparative word "more" (and by the same token, the word "less" as well).

An alternative hypothesis is proposed by this author in order to explain the Equalize and Compare difference. This hypothesis involves a combination of structural and linguistic factors and involves the two senses of the word "more" (or "less"), as identified earlier by Moore and Frye (1986). In an Equalize word problem, the "sequential" meaning of the word "more" is semantically synonymous with "equalize", as it refers to an increase or decrease within a quantity. A child understanding "more" in this particular linguistic framework, understands structurally, that an "action" relationship is taking place, as a comparison is made of the same quantity in two different states. In a Compare word problem, the "simultaneous" meaning of the word

"more" is semantically synonymous with "compare", as they refer to a comparison between two distinct quantities. A child understanding "more" in this particular linguistic framework, understands structurally, that a "static" relationship or no transformation is taking place.

Previous studies have been carried out concerning the difference between responses to Equalize and Compare questions utilizing discontinuous quantities only. A discontinuous quantity is a disjoint element (i.e. one which is made up of a number of separate parts). The opposite of this is a continuous quantity which is a joint element (i.e. one which is not made up of separate parts). However, word problems can be presented utilizing discontinuous quantities or continuous quantities.

Moreover, previous research does not distinguish between word problems involving discontinuous quantities and those involving continuous quantities. If quantitative comparisons are number-specific, then children should only be able to perform well with discontinuous quantities (which they can, for example, count), and not with continuous quantities (which are not number related). If quantitative comparisons are non-number-specific, then children should be able to perform well with both discontinuous and continuous quantities. This is important to examine in order to determine whether the solving of word problems by children involves a basic understanding of numerical concepts and a number-specific comparison. It is not clear why children are willing to use that understanding of numbers when solving Equalize questions and yet are reluctant to use that same understanding of numbers when solving Compare questions.

Children's understanding of numbers includes such methods as: one-to-one correspondence, counting, and addition and subtraction. Recent research, reviewed in Chapter 1, demonstrates that children have a basic understanding of numerical concepts (Bryant,1972; Gelman,1982; Fuson,1988; Desforges and Desforges,1980). Studies indicate that although

children might realize that two quantities are unequal, they have difficulty in comparing the relative composition of these two quantities, a task necessary to determining which quantity is greater and which smaller. A basic understanding of the above mentioned methods does not necessarily indicate a full comprehension of number knowledge. Rather, the use of such methods may indicate merely a general understanding of numerical concepts, without a specific grasp of quantitative comparison (Cowan,1987; Fuson, Secada, and Hall,1983; Michie,1984; Saxe,1977; Sophian,1987).

The aim of the experiments of this thesis is to investigate the Equalize and Compare difference, and subsequently to investigate the effect of discontinuous and continuous material, as well as the effect of different comparative terms, on the Equalize and Compare difference. Also, children's strategies will be observed in light of this difference.

Experiment 1 (Chapter 2) investigates whether children (ages 5-6 and 7-8) solve Equalize and Compare word problems involving continuous quantities in the same way that they solve word problems involving discontinuous quantities. Children's understanding of comparative terms in arithmetic problems and their ability to make use of graphical information in such problems is tested. Each child is shown a series of illustrations containing two numerical quantities. In half of these illustrations, numerical differences are represented continuously and in the other half, discontinuously. The children are individually tested on three types of quantities: continuous, discontinuous, and verbal discontinuous (where no pictorial quantity is presented), and are asked a question about each illustration. Half of the questions take the Equalize form, *"How much more does B need to have, to have the same as A?"*, and half take the Compare form, *"How many more does A have than B?"*. In all conditions, children perform better on Equalize-type questions than on Compare-type questions. Moreover, children perform best with discontinuous quantities, next best with

continuous quantities, and worst with verbal discontinuous quantities. Therefore, the Equalize and Compare difference applies to both types of graphical information (discontinuous or continuous). This indicates that the Equalize and Compare difference is not specific to number, but more fundamentally, to quantity in general. Such a finding has never before been reported.

These findings suggest that the ability of young children to deal with a mathematical problem is highly dependent on the way the problem itself is presented. It seems that children do not make use of basic numerical understanding when the problem requires the comparing of two continuous quantities.

Experiment 2 (Chapter 3) repeats Experiment 1 presenting the material on a BBC microcomputer and investigates strategies of quantitative comparison. Children (5-6 years) are presented with both discontinuous and continuous quantities. For both Equalize and Compare questions, two modes of expression are used: "more" or "less". Children make more correct responses in the Equalize-type question than in the Compare-type question in the "more" mode. All the children make mostly incorrect responses in the "less" mode. Children also make more correct responses in the Equalize-type question than in the Compare-type question with discontinuous material, but not, in this case, with continuous material. This finding is inconsistent with the results of Experiment 1. It suggests that the Equalize problem seems to be number-specific, whereas the Compare problem seems to be non-number-specific, as the difficulty children have with it applies to both discontinuous and continuous material. In making their quantity judgements for Equalize-type questions, but not Compare-type questions, children use strategies which involve one or more numerical methods. The strategies frequently involve numerical methods in the "more" mode, but never in the "less" mode. Children mostly use an Equivalence Strategy for the Compare-

type problems. This strategy entails choosing the choice stimuli that was the equivalent in height to one of the comparative pair. It seems that children have different ways of representing each type of problem.

Experiments 1 and 2 establish an Equalize and Compare difference. In pursuit of explaining this difference by the hypothesis which combines structural and linguistic factors, Experiments 3, 4, 5, and 6 implement structural manipulations. It is thought that by implementing these structural manipulations, the Equalize and Compare difference will not occur.

Experiment 3 (Chapter 4) investigates the child's understanding of part-whole relationships via equivalence. The experiment looks at children's preference for quantity equalization over quantity comparison and is concerned only with continuous quantities. This is done by setting up a perceptual bias in favour of quantity comparison over quantity equalization. This perceptual bias consists of lines, set up horizontally, dividing the larger stimulus of the comparative pair into two parts. The area above this line is denoted by A1; the area below this line is denoted by A2. The smaller standard of the comparative pair is another rectangle (B). The elements of the response set are vertical rectangles of same width, but different height either in ascending or descending order of height. There are three conditions: A1=B; A1≠B; Control. The control condition is the same as the experimental conditions with the exception that there is no line in the left-hand element of the stimulus set. A nested design is used whereby, within each condition, there are two types of questions and within each question there are two modes of expression, either "more than" or "less than". The stimulus set contains two elements. The response set contains five elements. The elements are vertical rectangles, all with the same base line.

There is no significant difference between the three conditions. Children perform better on Equalize-type questions than on Compare-type questions. Children perform significantly better in the "more" mode on

Equalize-type questions. Hence, performance on Compare questions is not affected by the part-whole manipulations which leads to the conclusion that part-whole relationships have nothing to do with the Equalize and Compare difference. Here, a structural hypothesis of children's difficulties with part-whole relationships is not supported.

As part-whole manipulations do not have any effect on the Equalize and Compare difference, Experiment 4 (Chapter 5) is then designed to look at the effects that other types of material could have on this type of problem. It investigates the differences between Equalize-type and Compare-type questions in another context, that of word problems. A word problem is a mathematical question disguised in a story. The terms "more" and "less" are substituted by the terms: "bigger" and "smaller", "taller" and "shorter", "richer" and "poorer". With each change in terminology, a story is narrated to the child. Discontinuous and continuous quantity are presented. Equalize questions are easier than Compare questions. Children perform better on Equalize-type questions which use the term "more" and those terms related to addition than to the term "less" and those terms related to subtraction. In Equalize-type questions, children perform better on those which entail discontinuous quantity than continuous quantity. These results replicate those of Experiment 2. Finally, children's performance is better in story-telling with continuous quantity than with discontinuous quantity. Story-telling, however, does not have any effect on the Equalize and Compare difference, indicating that the structural hypothesis cannot be explained in terms of this type of manipulation.

Experiment 5 (Chapter 6) uses a different tactic to investigate the Equalize and Compare difference. This difference is investigated by attempting to make the Equalize problem difficult. The experiment consists of spatial imagery manipulations which display two types of continuous graphical information in four possible permutations of identical and/or non-identical

quantity. As with Experiment 1, children perform better with questions which ask them to equalize amounts than with those which asked them to compare two quantities. Children also perform better with identical quantity than with non-identical quantity. Hence, the Equalize and Compare difference applies to both types of graphical information (same-type displays or different-type displays). It is concluded that the results of the spatial imagery manipulation do not support the structural hypothesis.

Experiment 6 (Chapter 7) further investigates the spatial imagery manipulation by presenting the two types of continuous graphical information in four possible permutations of identical and/or non-identical quantity (Experiment 5) in different-level displays. A different-level display constitutes the comparative and choice stimuli spatially arranged to begin at different base levels. A same level display constitutes the comparative and choice stimuli spatially arranged to begin at the same base level. It is thought that the Equalize and Compare difference would be extinguished by presenting the material on a different level, as children may be more prepared to be flexible in spatial representation and manipulation of spatial images. Children (6-8 years) are asked to identify which of the items of the choice stimuli is the correct choice for the difference between the two items of the comparative pair. There were five within-subject variables: Question Type (Equalize or Compare), Comparative Term ("more" or "less"), Display Type (same-level or different-level), Condition (same material or different material) and Material (bars or lines in the comparative pair).

There is neither an Equalize and Compare difference, nor an overall display effect. There is no difference between same-level and different-level displays. These results indicate that the structural hypothesis is not supported by this spatial imagery manipulation of the level of displays. However, the fact that there is not an Equalize and Compare difference suggests that children found Compare word problems easier to solve than in

previous experiments. This is explained by the fact that the comparative pair was not represented in the choice stimuli. Hence, the children are not able to use the Equivalence Strategy in the Compare questions in order to arrive at the correct solution of the problem. This supports, rather than negates, a structural hypothesis as it seems that children are able to solve Compare problems using a one-step process, rather than a multi-step process. They do not have to keep in mind a representation of the initial quantity and hence are able to solve the problem quite successfully.

In Chapter 8, a summary and integration of the experimental findings is presented. The theoretical importance of the experimental findings is discussed. There is a brief discussion of the implications of the present findings for teaching and suggestions for further experiments are offered. A major question emerges. Do children (5-8 years) use the same processes to solve both Equalize and Compare word problems? The results in this thesis suggest that children (5-8 years) use cognitive thought processes to solve Equalize word problems, but not Compare word problems. Children use only perceptual processes to solve Compare word problems. It seems that the representational systems that young children at first use to solve Equalize and Compare word problems are quite different. Hence, it seems that Equalize and Compare problems may exist in two different domains as far as young children are concerned.

CHAPTER ONE

1 CHILDREN'S MATHEMATICAL DEVELOPMENT - THEORETICAL BACKGROUND

1.1 Introduction

This investigation concerns the ways in which young children (ages 5 to 8) compare quantities and how they work out the difference between them. Although this may seem to involve a simple mathematical operation, this thesis proposes that the processes involved are complicated and crucial to the solving of addition and subtraction problems. Children have been shown to have difficulty comparing quantities, and in particular, unequal quantities. Yet, such a task is a fundamental part of mathematics.

Word problems are extensively used to test the mathematical abilities of children. They are stories which present a mathematical problem. One interesting empirical phenomenon concerns two types of word problems that involve the same mathematical operation. These are known as "Compare" and "Equalize" word problems.

"Compare" word problems engage the child in analyzing the composition of two distinct quantities to determine if there is a difference between these and, if so, what that difference is (i.e. 6-4=X). These problems involve a "static" relationship between two quantities. Static relations embody conceptual knowledge about comparisons involving sets of objects. An example of a Compare problem is the following:

Mary has 6 books. John has 4 books.
How many more books than John does Mary have?

"Equalize" problems involve a comparison between two quantities to discover the amount by which one of the quantities must be changed to make it equal to the other. Thus, a quantity transformation is achieved, involving an "action" relationship where addition or subtraction make one quantity equal to

the other (i.e. 4+X=6). Action relations embody actions that cause increases or decreases in some quantity. An example of an Equalize problem is the following:

Mary has 6 books. John has 4 books.
How many more books does John need,
to have as many books as Mary?

Previous studies have been carried out concerning the difficulty in responses to Compare questions, using discontinuous quantities only (where a discontinuous quantity is a whole made up of separate parts). This raises the question of whether the results obtained are linked to a number-specific context or to quantity in general. The use of continuous quantities (where a continuous quantity is a whole), could determine whether children make use of non-number-specific factors when solving word problems which require comparisons of two unequal quantities.

This investigation examines word problems using discontinuous quantities, as well as continuous quantities. It sets out to explore whether the patterns that are picked up on Compare word problems with discontinuous quantities are the same or different in Compare word problems with continuous quantities, and in Equalize problems with both discontinuous and continuous quantities. The difference between Equalize and Compare word problems using both discontinuous and continuous quantities is examined in order to find out whether there is a difference between the two, and if there is, whether this difference is specific to number. If there is a difference between Equalize and Compare word problems, and it is non-number-specific, then children's difficulty with quantitative comparisons is more fundamental.

On the one hand, Briars and Larkin (1984) conclude that the difference of word problems involve children's basic understanding of numerical concepts and a number-specific comparison. This hypothesis, however, does

not clarify why children are dependent on counting and on using that understanding of numbers in word problems, nor whether any difficulty encountered in word problem solving is specific to counting and number.

On the other hand, Riley, Greeno, and Heller (1983) hypothesized that children attempting to solve word problems requiring the comparing of quantities make use of a non-number specific process in reaching a solution. Hudson (1983) hypothesized the same for children attempting to solve Compare word problems. Riley, Greeno, and Heller's hypothesis concludes that children's differences in word problem solving are due to a misrepresentation of the word problem structure. Hudson's hypothesis concludes that children's difficulty with Compare word problems is due to a misinterpretation or inadequate comprehension of the comparative construction, thus linking his conclusion to a general linguistic factor. However, Riley, Greeno, and Heller's, as well as Hudson's, experimental results are limited to discontinuous quantities and the implications for non-number-specific factors lack concrete evidence.

In this chapter, a review of the literature on children's understanding of quantity in general will be presented. This is important as in order to know about comparing quantities, one has to know about the estimation of single quantities. Further on in this chapter, a review of the literature on children's understanding of quantitative comparisons within word problems, particularly that on Equalize and Compare word problems, will be discussed.

1.2 One-to-One Correspondence

Much research has been done on children's understanding of "quantity", which has yielded the conclusion that children as young as 3-years-old are capable of realizing that two quantities are unequal (Bryant,1972; Gelman,1982; Fuson,1988; Desforges and Desforges,1980).

Fundamental to addition and subtraction is the determination of which quantity is the "greater", and which the "lesser", of the two.

Children may use a variety of strategies to compare two or more sets of items. Some may be perceptual, such as comparing the lengths of two rows or comparing the relative density of the objects in distinct sets. Such strategies at times result in correct deductions; at other times, they can be misleading. More reliable strategies include establishing one-to-one correspondence between the objects in the different sets or counting these and later comparing them by means of addition or subtraction.

One-to-one correspondence is one of the more reliable methods that children may use in finding the difference between quantities. The method involves matching each object in one set with a counterpart in another set, and then determining how many objects remain after all matched objects have been cancelled. Two sets are equivalent in number when there is reciprocal correspondence between the objects in one set and the objects in the other. Two sets are not numerically equivalent when after cancelling out all objects in one set there is a remainder of objects not cancelled out in the other set.

For example, one-to-one correspondence of the objects in Figure 1 shows that row B is more numerous than row A. After matching and cancelling all objects in row A with those in row B, two objects are left over in row B. Therefore, B is the row with the greater number of objects and there is a difference of two between the rows.

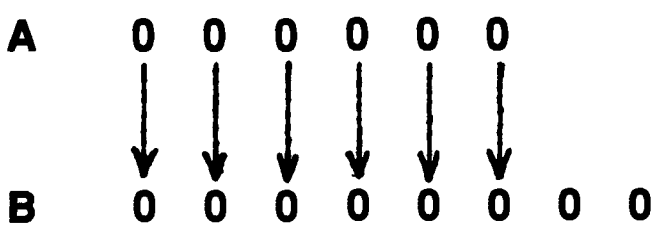


Figure 1. Illustrating one-to-one correspondence.

The understanding of this principle is essential to the understanding of numerical concepts, as no real understanding of the cardinal properties of numbers can exist without an understanding of the one-to-one correspondence principle.

One-to-one correspondence can involve spatial or temporal tasks. Spatial one-to-one correspondence involves making a quantitative comparison between two sets of objects laid out alongside each other. Temporal one-to-one correspondence involves the distribution of items into two or more sets. Most of the research carried out on children's understanding of one-to-one correspondence has focused on spatial tasks. Experimenters have concluded that children performing spatial one-to-one correspondence tasks are easily misled by perceptual notions such as length or density. This type of task has been found to be very difficult for children under seven. Evidence suggests that children have a better understanding of temporal one-to-one correspondence than of spatial one-to-one correspondence, at a very early age. Sharing behaviour, a type of temporal one-to-one correspondence, has been observed as early as 3 1/2-years of age (Desforges and Desforges,1980).

1.2.1 Spatial One-to-One Correspondence

Piaget and Szeminska (1952) were interested in determining the age at which children are able to use one-to-one correspondence successfully. Testing children of two different age groups: 4- to 5-years-old and 5- to 6-years-old, he instructed each child to construct a row made up of sweets identical to a model row previously constructed.

Children aged 4 to 5 typically proceeded by placing two sweets opposite each other, across from the sweets at the ends of the model row, and then filling in the gap with sweets without regard for correspondence to

sweets in the model row. This resulted in a row of the same length as the model row, but usually not containing the same number of sweets. The 5- to 6-year-olds, on the other hand, proceeded to make each row equal in both length and number of sweets to the model row. However, when Piaget proceeded to lengthen one row of sweets without actually changing the number of sweets, children of both age-groups claimed that the rows were no longer equal, and that the longer row had more sweets than the shorter row. The children adhered to their claim, regardless of whether they counted the sweets or not.

Piaget concluded that children under the age of roughly 8 cannot use one-to-one correspondence, but that children over the age of 8 can. He further postulated that even though the older children may make use of one-to-one correspondence they may not fully comprehend numerical concepts. Hence, he argued, a child may perform one-to-one correspondence and not necessarily understand what this represents mathematically.

Piaget and Szeminska (1952) and Gréco (1962) argued that children initially learn to count by rote; they merely parrot words without understanding their significance nor their purpose.

Gréco (1962) differentiated between,

"quotité" (=counting by rote)

vs.

"quantité" (=counting with understanding).

He postulated that 5- to 6-year-old children who do not understand one-to-one correspondence, and say, for example, that a longer row has more, even though they may be counting, do so because they understand numbers in a "quotité" fashion, and not in a "quantité" fashion. According to Gréco, children do not understand what a number really means. Despite the fact that children may be very good at counting, this does not mean that they understand numerical concepts.

It may indeed be the case that children over the age of 5 do not understand one-to-one correspondence. However, it may also be the case that these children are simply using a wrong cue. In the above-mentioned Piagetian task, length may be argued to be a conflicting perceptual cue and one which should not be used as the basis for judgement of children's understanding of one-to-one correspondence.

So, there was a need for studies where length would not be a conflicting cue. Further studies were conducted by Bryant (1972), who opposed Piaget's conclusions, and argued that children under the age of 5 can use one-to-one correspondence. Bryant affirms that these children simply do not realize that one-to-one correspondence is a more reliable strategy than length comparison. He presented 3- to 6-year-old children with various displays, each one containing two rows of counters (see Figure 2), and asked each child which row contained more. It is significant to note that all rows were composed of a large number of counters (some contained 19 and 20) so that young children could not easily solve the problem by counting.

Bryant's Counters Displays

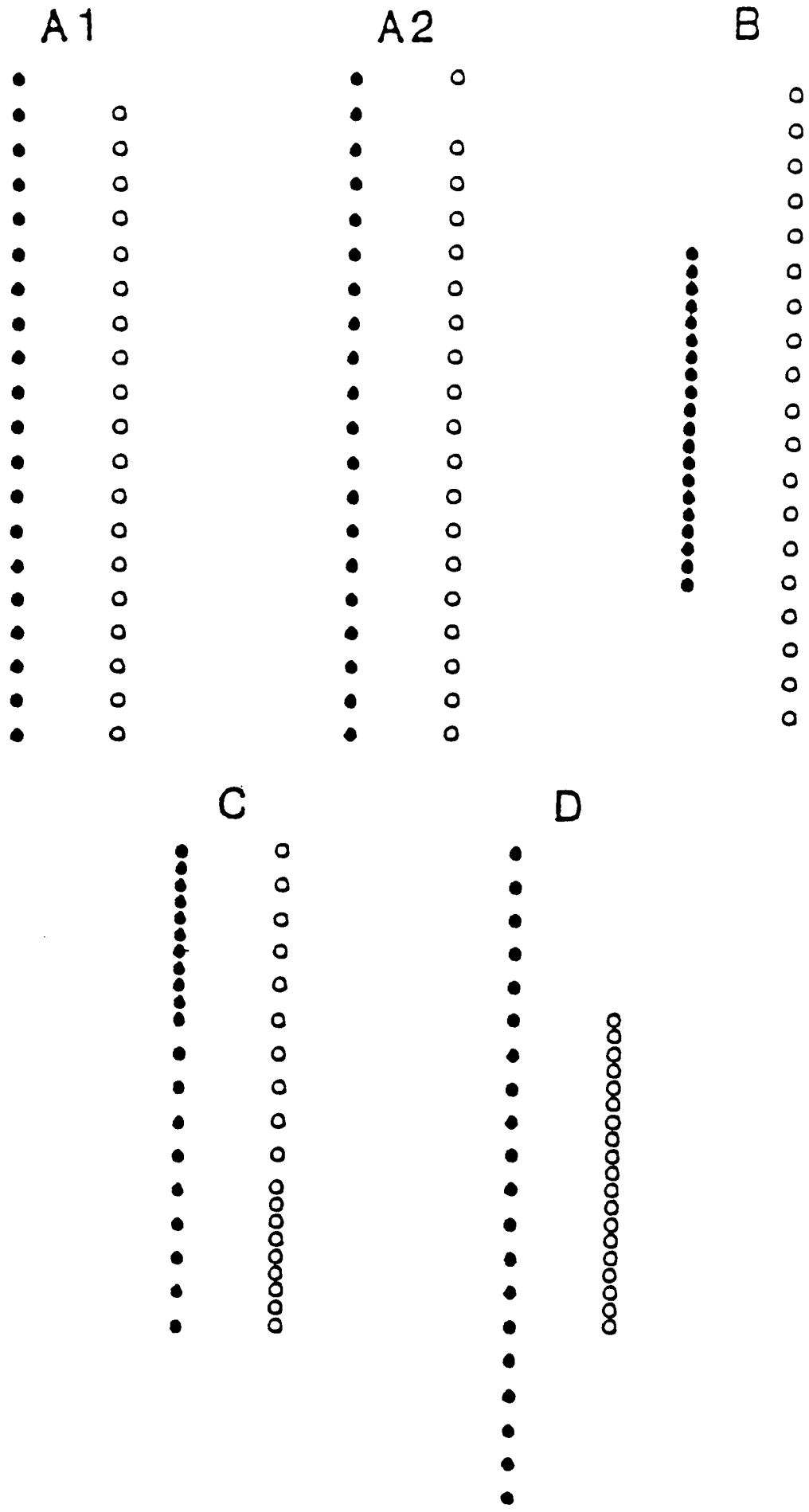


Figure 2. Examples of displays used by Bryant (1972) to assess understanding of one-to-one correspondence.

The A displays were arranged in such a way that one-to-one correspondence between the counters was obvious. The B and C displays were arranged so that one-to-one correspondence between the counters was not obvious. In the B displays, length was a misleading cue, and in the C displays neither length nor one-to-one correspondence was an obvious cue.

Children as young as 3-years-old, were consistently correct in judging the A displays; consistently incorrect in judging the B displays; and performed at chance level in judging the C displays. These results indicate that very young children can use one-to-one correspondence in judging relative numerosity, but they are also easily misled by length cues. The experiment supports the argument that young children's difficulties with quantitative comparisons stem from a distrust of one-to-one correspondence when conflicting perceptual cues are available. It seems to undermine Piaget's hypothesis that children under 5 are inherently unable to use one-to-one correspondence.

Bryant (1972) agrees that preschool children are easily misled by length cues and that length is indeed a conflicting perceptual cue. However, the length cue negates the one-to-one correspondence cue in Bryant's task in the presence of another perceptual cue, that of density. It may have been possible for children to understand both the one-to-one correspondence and length cues, in the presence of this other perceptual cue: density.

Another study, assessing young children's ability to use one-to-one correspondence when conflicting perceptual cues are present, was carried out by Brainerd (1973).

Brainerd presented six displays to children (ages 5 to 7), each one consisting of two rows of counters. Each display was in one of three modalities: lengths equal-number unequal; lengths unequal-number unequal; and lengths unequal-number equal (see Figure 3). The children were asked to judge the relative numerosity of each row in the display without counting.

Most of the kindergarten children (ages 5 to 6) and many of the first-grade children (ages 6 to 7) were unable to perform the task at all. Slightly under a third of the kindergarten children and slightly under half of the first graders were able to respond correctly to "lengths equal-number unequal" displays and to "lengths unequal-number unequal" displays. However, very few of these responded correctly to the "lengths unequal-number equal" displays.

Brainerd's Counters Displays

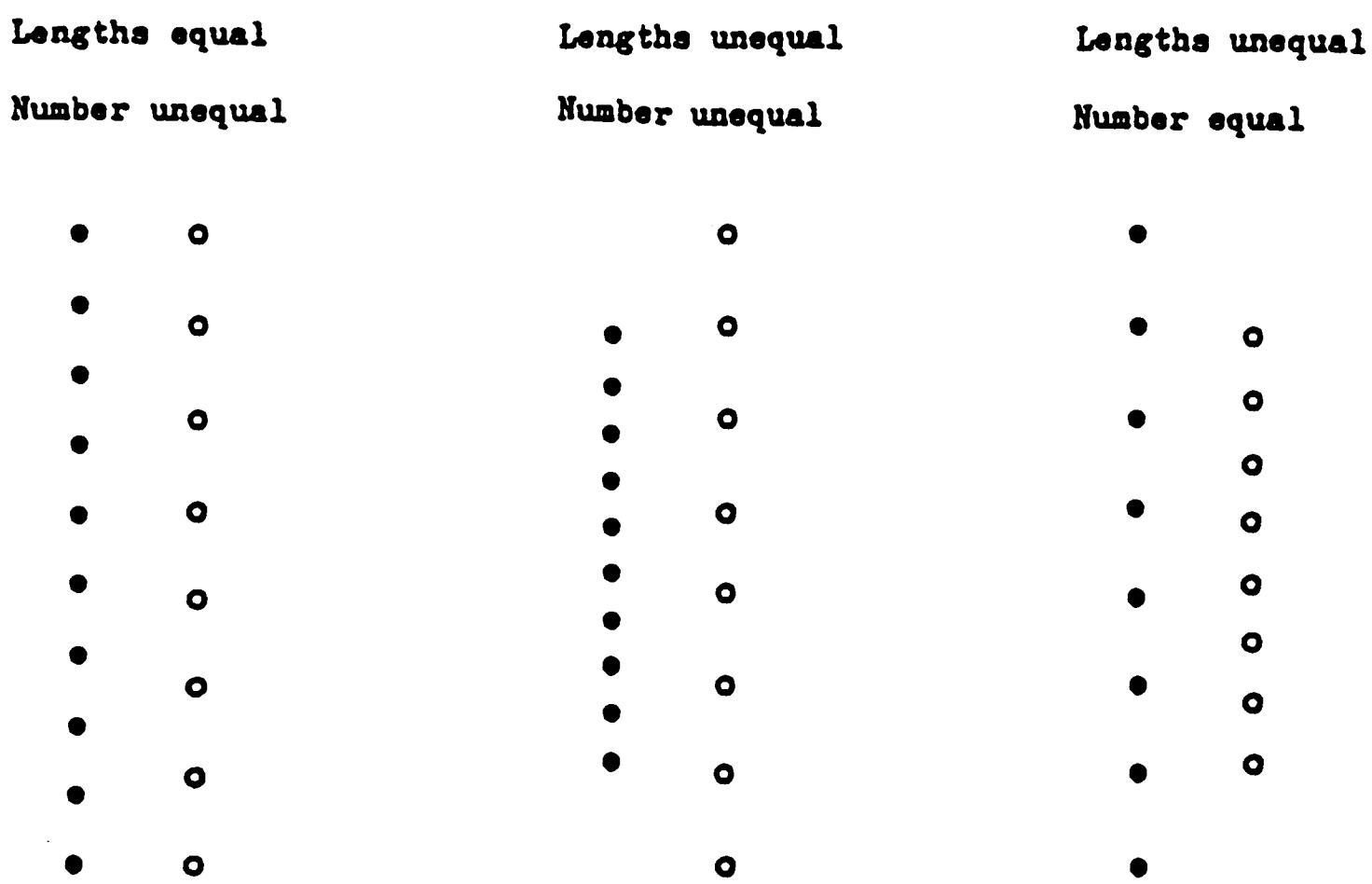


Figure 3. Examples of displays used by Brainerd (1973) to assess understanding of one-to-one correspondence.

These results lend support to Bryant's conclusion that young children (ages 5 to 7) tend to rely heavily on perceptual cues when comparing quantities, as was demonstrated by the majority of the children, regardless of age, being deceived by the lengths unequal-numbers equal display. The fact

that more than two-thirds of the children (ages 5 to 6) and more than half of the children (ages 6 to 7) were not able to solve the lengths equal-numbers unequal and lengths unequal-numbers unequal displays suggests that children are not able to use one-to-one correspondence at an early developmental stage, as Piaget had concluded, and seems to imply that the acquisition and understanding of this method is directly correlated with increasing age.

However, Cowan (1987) argued that most children fail Brainerd's (1973) one-to-one correspondence tasks because they fail to pair counters in the two rows correctly, not because they are unable to use one-to-one correspondence, as Brainerd concluded. Cowan proposed that the problem in previous research was that length and density were compounded. He proceeded to attribute Bryant's (1972) results (where children as young as 3 correctly judged quantities in perceptual correspondence displays) to children's use of relative density, rather than to children's use of one-to-one correspondence.

He noted that if the children's main difficulty in solving such tasks lay in their executing a pairing strategy, then their performance should be greatly improved by providing pairing cues. On the other hand, if the children were not able to perform a one-to-one correspondence task, then providing pairing cues should not alter their performance.

To test his argument, Cowan presented 5-year-old and 7-year-old children with the Brainerd displays, both in the original form and in a modified form where guidelines were added to make pairing easier (see Figure 4). Cowan's results duplicated those of Brainerd when the former made use of Brainerd's displays in their unaltered form. However, when he used modified displays, which provided pairing cues, the performance of the 7-year-olds significantly improved; the performance of the 5-year-olds, however, did not.

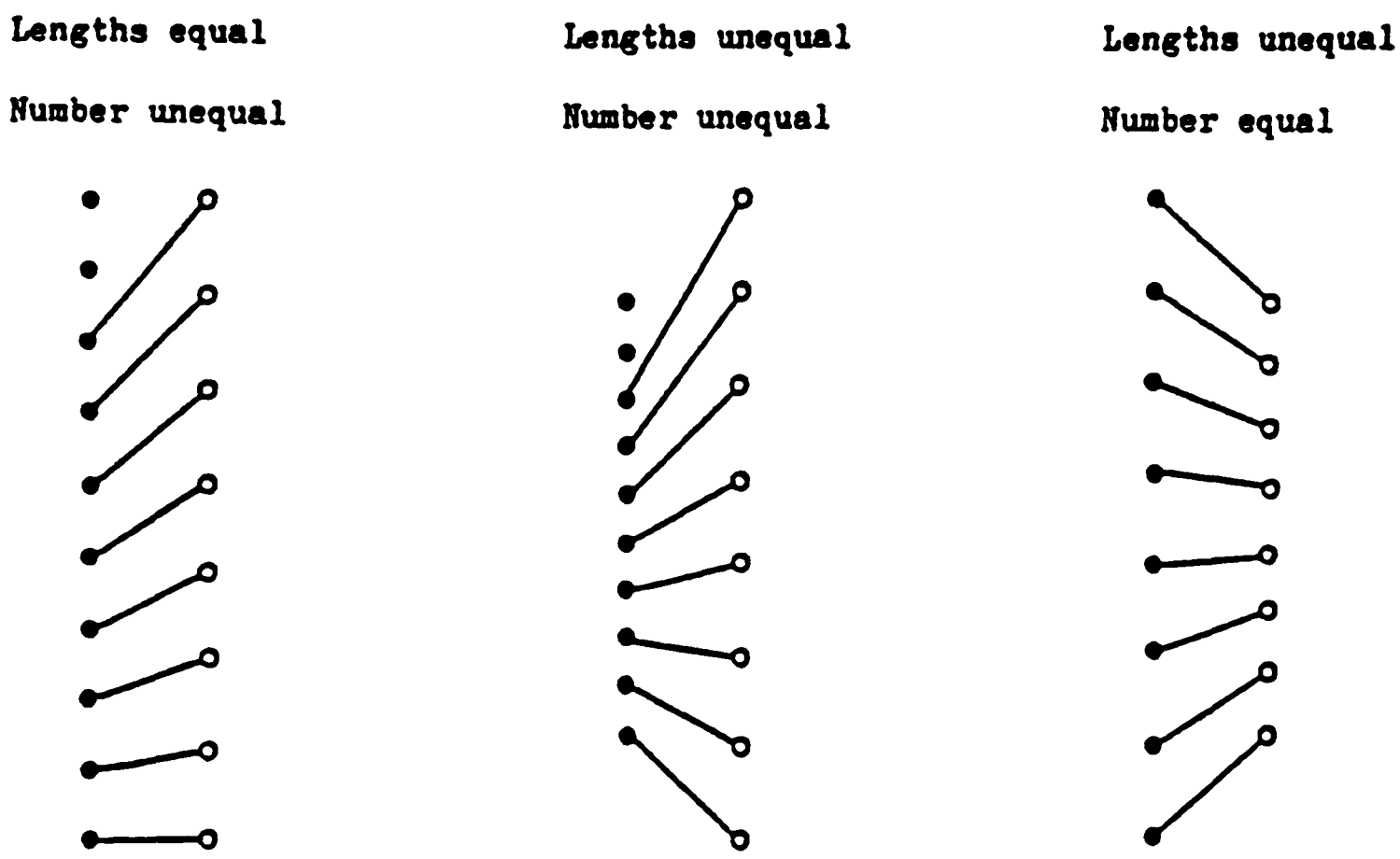


Figure 4. Examples of Cowan's (1987) modified Brainerd displays, with connecting guidelines.

These results indicate that children's difficulties with pairing items correctly are a contributing factor to children's performance on one-to-one correspondence tasks. In explaining the poor performance of the 5-year-olds as compared with that of the 7-year-olds, Cowan concluded that younger children, who are not helped by the pairing cues, rely on perceptual cues such as length or density to determine quantity because they are unable to use one-to-one correspondence. Older children, he argued, whose performance is significantly improved by providing pairing cues, may make use of these aids without fully comprehending why they work.

Cowan has concluded that older children do not understand one-to-one correspondence, despite the fact that they exhibit successful use of the strategy in his task. Guidelines, seem to have worked as compatible perceptual aids for the older children. It may be possible that they worked as conflicting perceptual cues for the younger children, hence their inability to use the one-to-one correspondence strategy to compare quantities successfully.

Gelman (1982), from a rather different point of view, was interested in assessing how the one-to-one correspondence principle could be made clear to preschoolers (3- and 4-year-olds) who understood numerical principles, as they demonstrated an ability to use a counting strategy, but could not conserve. She showed the children pairs of rows of objects in one-to-one correspondence that represented either equal or unequal quantities. A child was asked to count one of the rows and state its cardinal value. The child was next asked to count the other row and state its cardinal value. The child was finally asked if both rows had the same number. This final question dealing with equivalent judgments was not asked until the former two questions dealing with cardinality had been answered. This was done intentionally so as to demonstrate one-to-one correspondence in the definition of cardinal number by pointing out and making explicit the fact that displays which are (or are not) perceptual in one-to-one correspondence yield the same (or different) specific cardinal values when counted. The training included small set sizes of three to four items so that the young children could count them. The post-tests included large set sizes beyond the range of young children's counting capabilities.

Gelman's 3- and 4-year-old children were able to conserve on both small set sizes (four and five items) and on large set sizes (eight and ten items), after having been trained with the small set sizes. The training was done so as to make sure the children were able to count the items and come

up with the cardinal value. Since children of this age cannot accurately count sets containing more than four or five items, it was assumed that they used one-to-one correspondence in conserving the larger set sizes. Hence, according to Gelman, preschool children do use one-to-one correspondence when making judgments about equivalence.

Gelman's criterion for assessing a child's understanding of one-to-one correspondence is based on the child's ability to count. As long as the child assigns a unique count tag to each and every object in an array, the child is considered, according to Gelman, to exhibit an understanding of the one-to-one correspondence principle. This is regardless of the fact that the child may not recognize the equivalence relation between the number of count tags and the number of items in a display, as only one of these sets can visibly be seen; the other is represented in the mind of the child. Gelman concludes, on the basis of her above-mentioned results, that preschool children have an implicit knowledge of one-to-one correspondence when beginning to count, but one cannot assume that this implicit knowledge is readily accessible. Her results weakly substantiate her deduction that preschool children possess an implicit numerical and quantitative understanding.

1.2.2 Temporal One-to-One Correspondence

Temporal one-to-one correspondence involves the distribution of items into two or more sets. Thus, the study of sharing behaviour is one example of a temporal one-to-one correspondence task, which can provide important insight into children's understanding of numerical equality and one-to-one correspondence. It is also another way of showing how children arrive at comparing quantities. Unlike, spatial one-to-one correspondence, temporal one-to-one correspondence does not have the underlying limitations of

conflicting perceptual cues, such as length and density. The temporal version of one-to-one correspondence is fundamental to children's understanding of quantitative comparisons.

Desforges and Desforges (1980) looked at age differences in young children's (ages 3 1/2-6 1/2) sharing behaviour of equal and unequal quantities and at the relationship between set size and sharing errors. They also looked at children's sharing strategies for coping with remainders. In view of the findings that children perform better on number conservation tasks when set sizes are small (Gelman,1982), Desforges and Desforges varied the set sizes in this task to observe number-based strategies in sharing.

The experimenters asked children to share 5,6,9,10,11,15,20, and 30 "Polo" mints fairly between 2,3, and 5 dolls. Three different strategies were observed. Strategy A involved distributing the given mints one by one between the dolls until the set of mints was finished. Strategy B involved an attempt to divide the whole set into equal portions and allot one portion to each doll. Strategy C involved distributing the mints to the dolls in small groups of two or three rather than one at a time (e.g. *"two for this doll, two for that doll"*, etc.). Each of these strategies could be coded into two variants. Type 1 involved no apparent attempt to check or count as the sharing took place; children simply dealt the objects and apparently assumed that the dealing would lead to a fair solution. Type 2 involved the same manoeuvre as Type 1 but the process was accompanied by careful checking and counting with an overt reference to number (e.g. *"one for you, one for you, and one for you"* or *"this doll has two, she has two, she needs two"*).

Children used a non-counting/non-checking strategy when they divided the whole set of mints into equal portions and allotted one portion to each doll. Hence, these results paired Type 1 and Strategy B. This strategy was generally used by the younger children in the group. As children grew older,

they were less likely to use Type 1 (non-counting/non-checking) strategies and more likely to use Type 2 (numerical checking) strategies.

Use of Strategy C decreased significantly with age. Use of Strategy A, however, increased significantly with age. Desforges and Desforges kept records of the children's correct and incorrect answers including the way that the children dealt with remainders. The experimenters labelled a solution as correct if each doll received its proper share of the mints, or if, when a remainder was present, it was dealt with in some numerically appropriate way.

As the children's age increased, the tendency to remove the excess mints in order to equalize shares also increased. This behaviour was observed almost exclusively in the 5 1/2- to 6 1/2-year-olds. The youngest group of children (age 3 1/2- to 4 1/2) usually either asked for more mints to complete the share and make it even, or simply ignored the remainder and shared the mints without regard for equality.

These results seem to suggest that children by the age of 5 1/2 are quite competent at performing a task involving temporal one-to-one correspondence, and are certainly more successful than many researchers had found them to be at tasks involving spatial one-to-one correspondence (Piaget and Szeminska,1952; Brainerd,1973; Cowan,1987).

Three problems can be pinpointed in Desforges and Desforges' sharing experiment. (1) The negative results produced by the 3 1/2-year-old children may have been due to the fact that the procedure had too many trials (24: 8 set sizes x 3 divisors) and was too long. (2) Methodologically speaking, there may also be other ways of breaking up children's sharing. Perhaps other variants, other than Type 1 and Type 2 could have been considered. (3) A contestable point in their experiment concerns children's lack of sharing when failing to demonstrate appropriate distribution of mints among three dolls. They would distribute the mints in a random way, such as

"middle-1-1", instead of in a "1-1-1" way. These children demonstrated some confusion and probably lacked appropriate sharing skills. Desforges and Desforges provided no analyses for these observations.

It is possible, however, that the use of sharing as an indicator of children's understanding of one-to-one correspondence may lead to an overestimate of their understanding of this principle. Children may be able to use one-to-one correspondence when sharing, without necessarily understanding the quantitative significance of what they are doing. The one-to-one sharing procedure may simply have been learned by rote as a procedure that "works". If such is the case, then children should not be able to compare quantities effectively at an early developmental stage; nor should they do well in solving word problems which require such an ability.

To find out whether children do understand the quantitative significance of their use of one-to-one correspondence in sharing, Frydman and Bryant (1988) addressed the following questions:

(1) If objects have been shared equally between two people, and if children know how many items have been given to one person, will they know how many items have been given to the other?

(2) Can children adjust their sharing procedures when quantities have to be shared by, for example, giving single units to one person but pairs to another? In this case, the use of a "one for him, one for her" sharing procedure will not work because the second person would end up with twice as much as the first. Thus, if children have simply learned the one-to-one sharing procedure by rote, they will not be able to succeed in this task.

Frydman and Bryant designed a series of experiments aimed at assessing children's abilities to share and to make inferences about numbers on the basis of sharing. They asked 4-year-olds to share a set of 12 or 24 blocks amongst 2,3, or 4 dolls.

On 76% of the trials, the children performed in accordance with the one-to-one principle by distributing the blocks, one to each doll, until all blocks were distributed. The remaining 24% of responses comprised a variety of errors. The most common error was assigning a different number of blocks to the dolls at each turn. Another error was dividing the whole set of blocks into different piles at the start of the experiment, without reference to number. These errors involve a failure of temporal one-to-one correspondence, equivalent to placing the blocks of each pile on top of one another and using height to make them equal.

The 4-year-old children were, on the whole, able to share successfully on a one-to-one basis. On the other hand, the answer to the above question (1) appears to be negative: children, even after sharing sweets equally between two dolls and being told how many sweets had been given to one doll, were unable to state how many sweets had been given to the other doll. All of the children attempted to count the second doll's sweets, and when the sweets were hidden to prevent counting, only 10 out of 24 children gave the right answer. It can be concluded from these results that preschool children share by rote and do not understand the numerical significance of these repetitive actions.

However, Frydman and Bryant pointed out that children may know that sets that are shared on a one-to-one correspondence basis are equal (in other words, may understand the quantitative significance of sharing), but may not be able to express this knowledge in terms of numbers. They therefore carried out a second experiment on the sharing behaviour of 4- and 5-year-old children, in an attempt to answer the above-posed question (2). Here they varied the actual quantities that had to be shared out to two different recipients. Children were asked to distribute "pieces of chocolate" to dolls. These pieces of chocolate were blocks that a child could take apart to distribute in one unit singles or assemble into larger blocks of 2 or 3 units

stuck together. The children used four different strategies to complete the tasks of giving each doll a portion of the chocolate to be shared.

Strategy 1 consisted of randomly giving some blocks to one doll and some to the other, regardless of the numerical difference between the two dolls' portions and with no serious attempt to check how many units had been given to each doll.

Strategy 2 consisted of giving a certain number of blocks first to one doll and then to the second doll, after the number of portions to be given to each doll had been counted. The amount of units given to each doll was not verified.

Strategy 3 involved a simple quantity-based one-to-one correspondence. The blocks were distributed on a one-to-one basis, but all blocks were considered the same regardless of the number of units it had. Each doll was given one block at a time but no attention was paid to the difference between singles, doubles, or triples.

Strategy 4 (the correct strategy) involved a number-based one-to-one correspondence. Each time one doll was given a double or a triple block, the other one was given 2 or 3 single blocks, respectively. Alternatively, the child could first assemble the singles into doubles or triples and then carry on a simple one-to-one distribution.

Most of the children used strategy 3 for at least the first set of the two sets of trials. After the instructions were restated, two of the 24 children changed to using Strategy 4, the number-based one-to-one correspondence strategy, and performed correctly. Another child did the same, but only for the second trial. Two children adopted Strategy 2, the one involving counting, which helped them for one trial, but then they got lost at one stage of the counting process in the other trial and hence failed to distribute equivalent portions to the dolls. However, the great majority of these 4- and 5-year-old children persisted with the strategy with which they began.

The children, according to Frydman and Bryant, were finding it difficult to relate the outcome of Strategy 3, the one-to-one sharing strategy, to the outcome of Strategy 2, based on counting, which served to verify Strategy 3. Hence, the children found it very difficult to use their numerical knowledge in this type of task. Frydman and Bryant also concluded that these children did not have a good grasp of the sharing principle or of temporal one-to-one correspondence in general as they persisted in using the simple one-to-one sharing procedure throughout the experiment, as though by rote. Apparently, even though 4-year-olds share quite well, they do not have a strong grasp on the numerical or quantitative rationale of what they are doing. They fail to realize that distributing an equal number of blocks to each doll will not automatically result in both dolls getting equal portions. These children considered the one-to-one allocation of blocks to be the key factor in achieving equal portions, rather than considering the number of units that had actually been allocated.

The 5-year-olds performed much better than the 4-year-olds. Most used Strategy 4, the number-based one-to-one correspondence strategy, immediately. Those who did not, and who began to share the sets on the basis of Strategy 3 (disregarding the number of units making up a block), switched without hesitation to Strategy 4, the appropriate pattern, as soon as the mistake was pointed out by the experimenter.

Frydman and Bryant concluded that the understanding of the rationale behind sharing and the principle of one-to-one correspondence does indeed develop by around the age of 5. However, since most English children by the age of 5 have undergone some form of formal teaching, it is possible that this development is due to schooling.

Frydman and Bryant showed that children share well and way above chance with different set sizes between two to three recipients. However, the same underlying limitation that children may be adopting a procedure without

understanding could apply. The issue of whether children really understand the quantitative aspect of sharing remains.

Frydman and Bryant then decided to investigate whether 4-year-olds, whose results would not be confounded by schooling, could be helped to incorporate number into sharing. For this purpose, they designed an experiment in which 28 children were asked to share colour-cued blocks, each made up of one, two or three units. Numerosity was represented by colour: singles were one colour, doubles were another, and triples were of yet another colour.

The children were given a pre-test in a first session and a training and post-test in the second session. In the pre- and post-tests, the children had to deal out single block portions to one recipient and doubles or triples to the other. The blocks were red in colour.

In the training group, the blocks came in two different colours with different arrangements and therefore the significance varied between two groups. In the Experimental Group, the sets of single blocks included an equal number of blue and yellow blocks; the doubles were all made up of one blue block and one yellow block; and the triples had one blue block between the yellows or one yellow block between two blues. In the Control Group, the singles were the same as for the Experimental Group, but the doubles and triples were made up entirely of blue blocks or of yellow blocks.

The majority of the children in the Control Group displayed a quantity-based one-to-one correspondence strategy and only four out of 28 displayed a number-based one-to-one correspondence strategy. Two children unsuccessfully attempted to use counting in order to equalize the portion shared. The remaining four children distributed the sets without any obvious attempt at equality. In the Experimental Group, the colour cue led to a significant improvement in the performance of about seven of the children who had not done well in the task without the colour cue.

Frydman and Bryant concluded that the improvement of the children who had formerly not done well without the colour cues shows that the colour cues led them to take numerosity into account. They further concluded that pre-school children can share discontinuous material accurately, as long as the objects shared are all of the same number of units. However, the children's notion of the numerical significance of sharing is not well developed at this stage, as they were not able to share accurately when they had to deal out blocks of varying units. The 5-year-olds, on the other hand, were able to successfully incorporate numerical information into temporal one-to-one correspondence, thus demonstrating an understanding of quantity.

It is important to note the developmental change that transpires in children between the ages of 4 and 5. Frydman and Bryant suggest several possible alternative explanations for the 4-year-olds' difficulties with the task.

It could be, they postulate, that 4-year-olds' sharing is a rote procedure, learned by imitating others without any real understanding of how it works. Children know that the procedure involves "one for you, one for me", but they believe that the crucial factor is the successive allocation of objects to each recipient, without regard for the amounts of the objects being allocated.

A second explanation is that 4-year-olds initially do not realize that there are differences in the amounts of the objects being shared, but when they do they adjust their sharing procedure accordingly. The percentage of children who abandoned the one-to-one strategy when colour cues were used is so striking as to suggest that children do not always blindly follow the "one for you, one for me" routine.

Furthermore, it could also be that the children's basic understanding of one-to-one correspondence is more heavily influenced by perceptual than by numerical cues. The children know that there has to be a one-to-one match

when the quantities are shared, but their matching is guided by perceptual notions.

However, children do share. Five-year-old children share very well and have an understanding of temporal one-to-one correspondence; 4-year-olds share too, and can, in some cases, learn to incorporate numerical information into their behaviour when they share.

1.2.3 Conclusions about One-to-One Correspondence

Though it was previously suggested (Piaget and Szeminska,1952; Brainerd,1973; Cowan,1987) that children are not able to use one-to-one correspondence successfully, Bryant (1972) and Gelman (1982), claimed that children can successfully perform one-to-one correspondence, particularly when it is presented in temporal tasks (Desforges and Desforges,1980; Frydman and Bryant,1988). This suggests that children have a functional understanding of the numerical significance of one-to-one correspondence and therefore, of one of the basic rules of number. They are then able to implement one-to-one correspondence as a means of comparing two unequal quantities and finding the difference between them.

1.3 Counting

Counting is another method that children can use to compare numerical quantities. A child presented with two rows of objects can count the objects in each row and conclude that one row is more numerous than the other simply because it contains more objects than the other row.

Apart from its relevance to the comparison of two quantities, the understanding of counting would seem to be essential to a general understanding of numerical concepts. There is, however, considerable

controversy as to the extent to which children who perform the action of counting actually understand what counting means.

Counting works in a different way than does one-to-one correspondence. Like one-to-one correspondence, it is an important means for quantitative comparison. Unlike one-to-one correspondence, it is not a relative strategy whereby a direct comparison of two quantities can perceptually be made. Instead, it works as a measure whereby a child counts to arrive at the absolute amounts of the two quantities to be compared. Quantitative comparison may then be achieved by comparing the absolute numbers obtained via counting.

1.3.1 Children that Count: Their Understanding of Absolute Amount

It is essential to this investigation to review the evidence determining whether children do understand that successful counting can aid them at arriving at the absolute number of a quantity. It is also essential to review any further evidence of whether children understand that arriving at the absolute number of a quantity can help them to compare two unequal quantities and to work out the difference between them.

Gelman and Gallistel (1978) considered that counting in preschool children was important both in reflecting and in stimulating their understanding of numerical concepts. They affirm that children can count at the early age of two, argue that preschoolers do not totally lack appreciation of number invariance, and stress that children are not as ignorant of counting procedures as had been commonly thought.

Gelman and Gallistel analyzed children's counting processes and their understanding of absolute number. They concluded that children by the ages of 3 and 4 have some understanding of what they termed the four basic "how-to-count" principles, which are necessary to their understanding of

absolute number. The one-to-one principle stresses that children tag each object in a set just once; each item is given a number, or is counted, only one time. The stable-order principle indicates that numbers have to be produced always in the same order; it is always 1-2-3-4, and not 1-2-6-4 and later 1-2-7-8. The cardinal principle is the knowledge that after counting the objects in a set, the final number produced represents the number of objects in that set. Finally, the order-irrelevance principle states that the order in which the objects in a given set are counted does not affect the outcome of the counting process itself.

In their experiment, Gelman and Gallistel told 2-, 3-, and 4-year-old children to count single sets of objects ranging in number from 2 to 19. Nearly all the children, regardless of age, counted the small set sizes correctly. However, increases in set size resulted in an increase in overt counting and a tendency to err more frequently. The experimenters concluded that children understand the first three basic "how-to-count" principles of one-to-one, stable-order, and cardinality. In simpler terms, Gelman and Gallistel affirm that children understand what they are doing when they count.

However, it is not clear whether children have acquired a procedure which they do not understand, which is Piaget's view, or whether they do understand the principle of counting.

In another study, Gelman and Gallistel tested children's understanding of the order-irrelevance principle. Children (ages 3, 4, and 5) were asked to count objects in an unusual sequence. For example, the children were told to count the third object as 1, or the fourth object as 2, etc. Almost all of the 3- and 4-year-olds failed at this task; the 5-year-olds, on the other hand, performed successfully. Gelman and Gallistel concluded that children by the age of 5 understand the principle of order-irrelevance. The experiment, however, fails to support this conclusion. The results indicate that the 5-year-

olds are capable of successfully completing order-irrelevance tasks, not that they understand the principle of order-irrelevance.

Gelman and Meck (1983) sought an answer to why young children seem more successful at using counting principles, such as the order-irrelevance principle, when dealing with small numbers than when dealing with large numbers. They sought to explain whether children's mistakes with large set sizes were due to a conceptual difference or to carelessness. They suggested that this may be due to children having production difficulties in counting large quantities, even though they understand the basic "how-to-count" principles. The experimenters attempted to remove production difficulties by asking 3- and 4-year-old children not to count themselves, but rather to evaluate the counting of puppets, which sometimes conformed to all the counting principles, and sometimes violated one or more of them. The puppets counted sets containing 5 or 7 objects. The children were asked to judge if the puppets had counted correctly or not. An incorrect form of counting would have the puppet skipping over an object or double-counting an object. A second type of error - a pseudoerror - would have the puppet begin counting from the middle of the row of objects, for example, or skip an item in the middle of the row and at the end return to count it last.

Children's performance when they judged the counting performance of puppets was not affected by the size of the sets as it was when they did the counting themselves. They were actually very good at judging the counting performance of puppets. Gelman and Meck concluded that children's knowledge of the counting principles of one-to-one and cardinality may be underestimated as a result of their production difficulties and carelessness with large sets. They believe that children have an understanding of the counting principles at the moment they begin to count, and increasing age merely hones their skill at using these principles.

However, their experiment does not lend support to this theory as the children might have known the correct procedure and, therefore, have been able to decide when the puppet had counted incorrectly, without understanding the underlying principles. The design of the experiment was not such as to prove the children's understanding of these principles. If cardinality would have been assessed in its true meaning, and not in the Gelman and Meck sense, perhaps children would have used counting as an appropriate measure of quantitative comparison.

On the other hand, Schaeffer, Eggleston, and Scott (1974) found that 3- and 4-year-olds gave evidence of not comprehending the cardinal rule when counting an array containing more than 5 objects. Although the children could recite the numbers as they counted the objects, when asked for the total number of objects in the array, they often reported a number other than the final number in their count sequence. The experimenters concluded that the last number recited did not represent for the children the cardinal value of the array, and hence, the children had no understanding of cardinality.

Also assessing children's understanding of the basic counting principles, Baroody (1984d) argued that Gelman and Gallistel's test of the order-irrelevance principle (which states that the order in which the objects in a given set are counted does not affect the outcome of the counting itself) did not measure what it was intended to measure. On the one hand, children might understand the order-irrelevance principle but fail at the task of making the same object the "one", the "two", etc., because this rather unusual procedure is bewildering to them. On the other hand, children might be able to change the order in which they count, without realizing that this is irrelevant to the number obtained.

Baroody presented preschool children (ages 5 and 6) with a row of 8 objects, which they counted from left to right. After being asked, "How many

are there?", the children were repeated the number obtained in the counting sequence, "We got N counting this way" (i.e. where N is the number of objects counted), then asked how many items there would be if they counted in the other direction (from right to left). Most of the 5-year-old children, 55%, did not respond with the same number. However, most of the 6-year-old children, 87%, completed the task successfully. Baroody concluded that the younger children did not understand the order-irrelevance principle.

Gelman, Meck, and Merkin (1986) argued that the mere fact that the children in Baroody's experiment were asked the "how many are there" question twice, might have given the children the idea that they were expected to respond differently on the second occasion. So as to eliminate this possibly confounding factor, Gelman, Meck, and Merkin designed an experiment in which 4-year-old children were given a task similar to Baroody's, and added two conditions. Children were asked, in one condition, to count left to right 3 times (rather than just once). In the other, more crucial, condition, children were asked how much they would get counting the other way. The children performed significantly better under these latter conditions, than under Baroody's. More children responded with the correct number.

Gelman, Meck, and Merkin concluded that preschoolers do understand the order-irrelevance principle, and that Baroody's design biases the children towards changing their answer. On the other hand, it is possible that the children were successful in Gelman, Meck, and Merkin's study because they had just counted up to a particular number, which they proceeded to reiterate in response to the experimenters' question, and not because they really understood the order-irrelevance principle.

The importance of counting and the "how-to-count" principles rests on the concept of absolute number. However, a child may know how to use counting to come up with an absolute number, but may not know how to

make further use of it as an intervening measure for comparing two sets of unequal quantities.

1.3.2 Children's Use of Counting to Compare Quantities

If children know how-to-count, and if they understand the basic principle of cardinality, then they should use this knowledge and skill to compare two or more sets of objects. However, counting is one aspect in the numerical domain; knowing how to use counting to make a comparison is another aspect in the numerical domain. One aspect does not guarantee the other.

Michie (1984) addressed the problem of why preschool children, even though they can count, frequently do not use counting to compare quantities. She was particularly interested in finding out if children's judgments would be influenced by counting, even if other cues to number, such as length and density, were present.

Michie investigated 3- and 4-year-old children's judgments concerning which of the two arrays contained more items. She asked them to make the comparison under different conditions. The use of perceptual cues of length and density was assessed by means of comparing arrays in which they conflicted (and therefore provided contradictory information) with arrays in which they did not conflict. The understanding and use of counting were examined by comparing the accuracy of children's judgments when children were and were not asked to count the arrays, as previous pilot studies had shown that preschool children generally do not count unless asked to do so. Finally, the use of the third potential source of information about numerosity, subitizing (i.e. the abstraction of the actual number of items in a display by directly perceiving the numbers), was investigated by comparing judgments of

small numbers within the children's subitizing range with judgments of numbers too large to subitize.

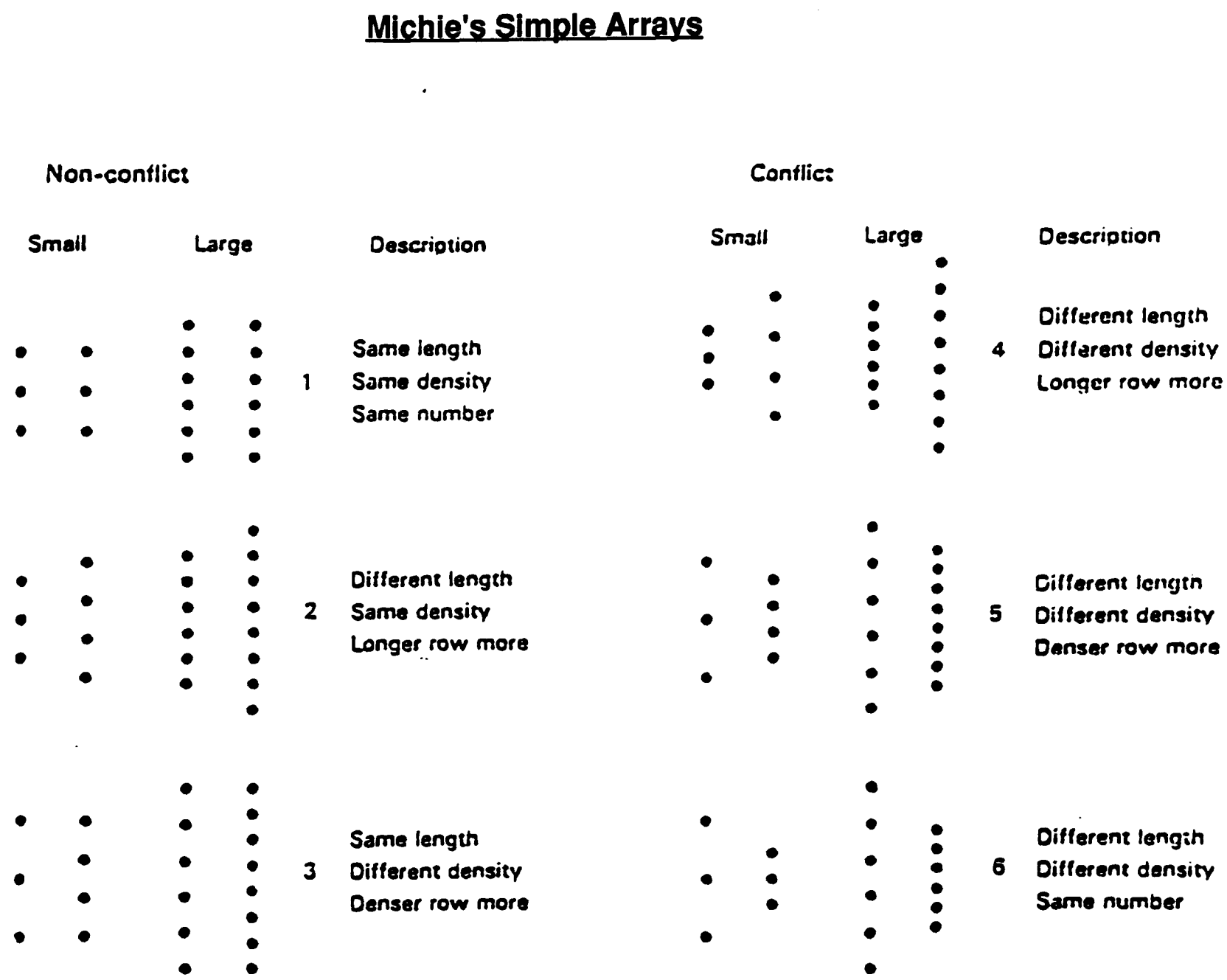


Figure 5. Examples of Michie's (1984) simple arrays for assessing children's use of spontaneous counting.

Results indicated that both 3- and 4-year-olds were more accurate on non-conflict than on conflict arrays. Greater accuracy was also found with small numbers than with large numbers. The great majority of the errors was due to children saying that the longer row had more, when in fact, it had less than or an equal number of items as the shorter row. The 3-year-olds made

463 length errors and only 50 density errors and the 4-year-olds made 307 length errors and only 37 density errors from a total of 1080 trials.

From these results, Michie concluded that preschool children used information obtained through counting and subitizing small numbers, though they preferred length cues, when making relative number judgments. Length appears to be only one of the many cues by which children judge numbers. This conclusion contrasts with previous conclusions that children base their number judgments exclusively on length (Piaget and Szeminska,1952; Brainerd,1973). Furthermore, Michie found that children did not count spontaneously in her task.

Michie did not discuss the fact that in her experiment, the two arrays were presented simultaneously. It has previously been suggested (Wilkening, Levin, and Druyan, 1987) that children may be more prepared to count with a successive presentation rather than with a simultaneous presentation.

Michie suggests possible reasons for children's infrequent use of counting. This accounts for their infrequent use of counting in comparing two unequal quantities. She considers that it may be due to a lack of knowledge about the value of counting as a cue to number, rather than to a lack of understanding of what counting really means. In simpler terms, Michie states that children may know that counting provides numerical information, but may not realize just how reliable this information is in working out the difference between two quantities, especially when compared with information from other sources.

If children simply fail to count spontaneously in such spatial tasks, but are capable, once instructed, of using the numerical information thus obtained appropriately, then one would expect that when children count the items in two arrays that are to be compared, their judgments of "which is more" should be correct.

Fuson, Secada, and Hall (1983) investigated children's ability to use one-to-one correspondence in making equivalence judgments in Piagetian conservation of number tasks, and the extent to which they use these techniques spontaneously. The use of matching to establish one-to-one correspondence (and hence numerical equivalence) between objects in two arrays can be very effective. Fuson et al. (1983) cited Brainerd (1973), who suggested that children rarely use matching to establish one-to-one correspondence, and that, when they do use it, they use it ineffectively. Brainerd, however, did not address the question of whether young children are able to use the correspondence information derived from matching to make equivalence judgments in comparing two unequal quantities. Therefore, Fuson et al. claim that Brainerd's study did not provide sufficiently clear evidence to distinguish between the possibilities that:

(a) children do not use one-to-one correspondence spontaneously (similar to Michie's (1984) claim that children do not count spontaneously when making relative number judgments, though they can do so);

(b) they cannot apply one-to-one correspondence effectively; or

(c) they cannot use the information derived from one-to-one correspondence, as they do not understand what it means.

They contrast this evidence with Whiteman and Peisach's (1970) findings that children's performance on a Piagetian number-conservation task was improved by a cueing procedure where corresponding pairs were connected by lines. As previously mentioned in Section 1.2.1 of this chapter, Cowan (1987) obtained similar results to those of Whiteman and Peisach (1970).

Little clear evidence has been obtained as to whether any children use matching and counting simultaneously and spontaneously to make numerical equivalence judgments nor to how young they are when they do so. Fuson et al. (1983) carried out an experiment with 4- to 5-year-old children, aimed at assessing the effectiveness of counting and matching in a Piagetian number

conservation task involving judgments of numerical equivalence. There was a standard conservation condition, which was set within a context of "feeding the animals at the zoo".

Seven toy animals were lined up in a row in front of a child. A row of peanuts was placed in front of the animals so that each animal had a peanut. The child was asked an initial equivalence question. The child was then told that the animals had to move to new places and helped the experimenter in spreading the animals apart so as to make a longer row. The child was then asked a conservation question, *"Are there more animals than peanuts, the same number of animals and peanuts, or more peanuts than animals?"*. The child proceeded to be asked, *"How do you know?"*. The same procedure was used with the count condition, with the exception that after the animals had been spread apart, the child was asked to count each row, with assistance if needed. To ensure attention to the count information and memory for both counts, the child was then asked, *"How many animals are there?"* and *"How many peanuts are there?"*. The child was reminded of the answer obtained, if necessary. Then the conservation question and the justification question were asked. The match condition was the same as the standard condition with the exception that before spreading the animals apart, the child was helped to place a string connecting each animal to a peanut. After the transformation the child was asked three times which peanut went with which animal. If the child gave an incorrect response for any of the three animals, the child was told to "look at the string" and was asked the same question about another animal. Again, the conservation question and the justification question were asked.

The hypothesis was that children would initiate matching and counting in such tasks prior to the acquisition of Piagetian conservation if indeed matching and counting experiences do contribute to the acquisition of Piagetian conservation.

Fuson et al. (1983) reported that the only visible behaviours observed before the children made equivalence judgments were counting (11 out of 16 children responded correctly in the count condition) and matching (12 out of 15 children responded correctly in the match condition). Hence, the responses were consistent with the view that most children move from the use of a perceptually based response strategy to the use of counting or of matching with some Piagetian justifications before actually conserving (only 2 out of 14 children responded correctly in the standard condition). Fuson et al. (1983) also suggested that the child who moves from incorrect perceptually based responses directly to Piagetian reasoning responses may very well be the exception rather than the rule. It is important to note that these interpretations were also based on the finding that perceptually based justifications were associated with incorrect responses, whereas Piagetian justifications were always associated with correct equivalence judgments.

Fuson et al. (1983) confirmed the common observation that most children aged 4;6 - 5;6 rely heavily on perceptual features of arrays (e.g. length) in making numerical equivalence or nonequivalence judgments of such arrays. However, when children were cued to use counting and/or matching in making their judgments, correct equivalence judgments when working out the difference between two unequal quantities were made. When using this information, children exhibited little difficulty, if any, in executing the counting and/or matching procedures and then making the correct equivalence judgment, regardless of the fact that this information was in direct conflict with the perceptually based information. Few children had any difficulty executing the counting or matching strategies. What these children did not do, with few exceptions, was to initiate such strategies spontaneously. This finding is similar to Michie's (1984) finding (mentioned earlier) that children can use counting rather than perceptual cues in relative number judgments, if they are given feedback about its reliability. Children did not

use these two procedures spontaneously, but once cued to use them, they performed well. It was the older children who did demonstrate considerable spontaneous use of both counting and matching in the number conservation task leading up to correct conservation judgments in comparing two unequal quantities. Fuson et al. (1983) stress the importance of a developmental sequence which involves the use of counting and/or matching, as well as awareness of the unreliability of perceptually based responding, in successful performance of Piagetian conservation tasks. It is through the experience of using counting and/or matching that children learn the principles underlying the counting system and thus, transforms developmentally children's understanding of number. Fuson et al.'s conclusions may be right. Children originally learn to count as a rote procedure with little understanding of what it means.

Cowan (1987) was interested in Michie's (1984) and Fuson et al.'s (1983) conclusion that children can count two sets of items correctly, but fail to make an accurate relative number judgment when working out the difference between two unequal quantities. He wanted to see whether children who were allowed to count the items in given rows would make relative number judgments that were consistent with their counting but inconsistent with relative length. He conducted four experiments with children between the ages of 3 and 7.

There were two phases to his experiment where the children judged small and large number displays, length consistent and length inconsistent displays, as well as equal number displays and unequal number displays. Displays were of 3, 4, 8, 9, 15, and 16 dots. In the first phase, children's counting of objects in single rows was assessed in a context that tested their consistency and established lower limits on their counting range. In the second phase, children's counting of objects in two rows was assessed and counting of the items in each row was done by either the children themselves

or the experimenter. Results of the counts of each row were repeated to reduce the likelihood of children forgetting how many items there were in the first row after they or the experimenter had counted the second row. Then a relative number judgment was elicited.

Results from the first phase demonstrated that 7 out of 65 children counted the rows of 8 and 9 with assistance, but failed on the rows of 15 and 16; 12 children managed the rows of 8 and 9 without assistance, but failed on the rows of 15 and 16. Eight children succeeded in counting the rows of 15 and 16, 2 of those 8 managed without assistance. The remaining 25 children of the 65 had to have their sessions terminated due to a lack of interest or to an inability to count the rows of 3 and 4 dots.

Results from the second phase demonstrated that, for the numerosity displays, children made fewer correct judgments when comparing two unequal quantities on the larger number displays than on the small number displays whether or not the judgment suggested by relative length was consistent. For the length consistency displays, children were more likely to make correct relative judgments of displays. For the length inconsistency displays of unequal numbers, children were more likely to judge the lengths equal displays correctly more than the lengths unequal displays. For the equal vs. unequal numbers displays, children were more successful in judging equal number displays than unequal number displays.

When judging the length inconsistent displays, 80% of the errors were length-consistent. On the numbers unequal-length equal displays, 74% of the errors were length-consistent. On the numbers unequal-lengths unequal displays, 82% of the errors were length-consistent. Finally, on the numbers equal-lengths unequal displays, 83% of the errors were length-consistent.

The main results of the second phase indicated that children did make relative number judgments inconsistent with counting when working out the difference between two unequal quantities. Children only made counting

consistent judgments more frequently on small number displays than on large number displays. Also, equal length displays were judged to be equal in number and the longer row was judged as having the bigger number in unequal length displays.

Cowan concludes, from his results, that counting is context-based by a counting based system for quantitative comparison or a pre-existing perceptual feature based system. Cowan further states that children, when using counting to determine the relative number of two sets do need to know how to count one set and stop, storing the last number counted, and then count the second set and stop, storing the last number counted. Several children counted both sets without pausing. Hence, children seemed to be lacking an understanding of numbers that allows them to judge correctly whether two sets have or have not the same number. With regards to counting being a pre-existing perceptual feature based system, Cowan, like Michie (1984), states that with small numbers, children may have two systems, subitizing and counting, which yield the same relative number judgment. On larger number displays, children may have counting and estimating (which may yield discrepant judgments).

Hence, Cowan demonstrates that young children are more likely to make accurate relative number judgments when comparing two unequal quantities after counting than after simply looking at large number displays or attempting to apply one-to-one correspondence to such displays. The difference in performance according to whether children are allowed or not allowed to count is striking. It seems, from Cowan's results, that children usually do not spontaneously count to compare two sets.

Cowan, however, does not explain the set size effect. Counting, is expected, to be affected by set size, whether it be spontaneous or not. He did not give children, the possibility of re-counting the sets, so as to verify cardinality. The experimenter reminded the children of the two numbers

before asking them to judge relative numerosity. Perhaps, this was not enough for children to perform an appropriate comparison between the sets, especially when the numerical information conflicted with the perceptual information. It is unclear whether Cowan regards children's incorrect judgments as conceptual or procedural difficulties.

Another piece of evidence of children's reluctance to use counting when comparing sets is a study by Saxe (1977). He regarded counting as an essential element for comparing sets and working out the difference between them. He gave several number tasks to 3-, 4-, and 7-year-olds. These included a test of rote counting ability, where the children were asked to count as far as they could; an object placement task, where the children were presented with a linear array of nine beads and were then asked to "put out just the same number" from an identical set of 15 beads; a drawing task, where the children were presented with a linear array of nine cardboard circles and were asked to "draw just the same number" of circles on a piece of paper as there were in the model array; a remote placement task, where the experimenter gave a puppet on the floor nine small toy animals to "eat" and then instructed the child to give a puppet on a table "just as many" from an additional set of 14 small toy animals; and a comparison task, in which children were presented with two unequal linear arrays of nine horses and 11 pigs, so that the endpoints of two rows were aligned in spatial comparison. The children were asked, *"Are there just the same number of pigs and horses, or does one have more?"*. In all tasks if children did not count spontaneously, they were asked: *"Would counting help?"*.

It is quite difficult to establish what actually is happening in Saxe's research. At times, the theory cannot be distinguished from the results.

Saxe noted an extreme reluctance to count in order to compare the two arrays on the part of the 3- and 4-year-olds. The remote object placement task elicited the least amount of counting. As this task contained

no other cues, such as perceptual or correspondence cues, it was assumed that counting would be most necessary, but it was not. The 7-year-olds, on the other hand, did count to compare both arrays. Saxe concluded, and subsequently expanded into a complete theory (Saxe,1979), that with increasing age, children progressed from what he termed "prequantitative counting strategies" to "quantitative counting strategies". Quantitative strategies were hardly ever used by the 3-year-olds, and were practically generalized to all tasks by the 7-year-olds. In his view, only children who use counting in making numerical comparisons between two unequal quantities are, in effect, counting "quantitatively". Until that stage, their counting is "prequantitative". By 4-years-old, 43%-to-69% of the children, varied in their use of quantitative strategies.

In prequantitative counting, Level 1, counting is not used as a means to produce numerical comparisons or reproductions of arrays (where the child makes an approximate copy of a model). When while performing a quantitative task, the child is asked, *"Would counting help?"*, the child counts but does not adjust the copy on this basis. For comparisons, the child initially counts only one set or both sets continuously as though they were one. At a slightly more advanced stage the child counts both sets separately, but does not use counting as a basis of comparison to work out the difference between two unequal quantities.

At a transitional stage between prequantitative and quantitative counting, numerical comparisons based on counting and spatial perceptions (i.e. length) compete with one another. The child either modifies counting to make it conform to spatial perceptions or fluctuates between evaluations based on counting and spatial perceptions. A third of the 4-year-olds were rated as being at this transitional stage. The children counted both arrays, but were unable to articulate their assessment of relative numerosity based on counting with that based on spatial cues. Thus, the child would establish

the number of items in both arrays, but disregarded the result of counting and claimed that they were of the same number. Other children seemed to overtly miscount the second array so as to end with the same number as in the first array.

In quantitative counting, Level 2, counting is used to produce numerical comparisons and reproductions of arrays. The use of both these strategies is closely related to general counting accuracy and to age, with most 3-year-olds using prequantitative counting strategies, and most 4-year-olds and all 7-year-olds using quantitative strategies.

As children come to understand that in order to determine the numerosity of an array, the number names must be applied in a serial one-to-one correspondence to objects, they begin to count more accurately. From this, Saxe concluded that as counting accuracy and counting strategy develop simultaneously, they must be ruled by the same underlying cognitive processes. This is due to a greater understanding of the logic of the counting system and to the use of quantitative strategies, as the children come to understand that counting is a way of extracting numerical information from arrays when comparing two unequal quantities.

Saxe's theory has mostly been supported by experiments which involve sets with large numbers of objects (more than nine objects) which are not easy for a child to estimate without counting. It deals mostly with children's manipulation of numerical information which is not "concrete" or "perceptually salient". It is important to note that, as he himself points out, different results might be obtained when sets with few objects are used. He also does not distinguish between spontaneous counting and non-spontaneous counting. It is then validly assumed that the children's use of counting was triggered by the experimenter's cue.

A more convincing test of children's understanding of number as a measure was devised by Sophian (1987). She investigated this type of

variable by examining which would be the most sensitive measure of children's knowledge about counting and numerical comparisons: counting tasks or judgements about counting tasks. Sophian (1987) was particularly interested in preschool children's judgements of other people's counting when comparing two sets. Gelman and Meck (1983) had established with their puppet experiments that preschool children are best at making judgements about those aspects of counting that are reflected in their own counting. Therefore, Sophian had previous evidence that young children would be good at making judgements of other people's counting of single sets since they themselves engage in this sort of counting without difficulty, but that they would encounter difficulty when making judgements about the use of counting to compare two sets, as they have difficulty with this sort of counting (Briars and Siegler,1984; Gelman and Meck,1983; Gelman, Meck, and Merkin,1986). Sophian wanted to find out why this was the case.

Sophian observed 3- and 4-year-olds making judgements about a puppet's counting two distinct sets of objects. In one task, the puppet had to say how many objects there were altogether, and in the other task, the puppet had to compare the number in each set. For both tasks the puppet counted all objects as one set, which would be the correct strategy for the how-many task, or counted both sets separately, which would be the correct strategy for the compare-sets task.

Sophian found that preschoolers were quite good at determining "how many" in a count, but that this was not the case for determining "compare sets" in a count. Children were good judging whether counting was done correctly or incorrectly when a "how many" task was concerned, but were not above chance level when a "compare sets" task was concerned. Sophian's finding supports the claim that preschoolers do not yet understand the way in which counting can be used to compare sets in order to work out the

difference between them, and therefore concludes that they do not have a real understanding of cardinality when they count.

Sophian's study provides evidence demonstrating that preschool children have a limited understanding of how to use counting to solve different types of problem. If children do not understand that counting is a means of quantification, they will have extreme difficulty not only in using counting to compare two sets, but also in using it for any meaningful mathematical computation (Saxe,1977; Schaeffer et al.,1974; Sophian,1987). This has important implications both developmentally and educationally. Developmentally, it supports the view that children do not have an innate conceptual understanding of what they are doing, but only gradually acquire that understanding presumably through their counting experiences. From an educational point of view, limitations on children's understanding of counting are fundamental because mathematical instruction may presuppose knowledge of counting strategies and skills that children do not necessarily have when they start school. It is then up to educators to provide children with instruction in these crucial aspects of counting, either as part of the standard mathematical curriculum or through a remedial program for children who are at risk of mathematical failure, as knowing how to use counting to compare two sets of unequal quantities and working out the difference between them is highly relevant to an understanding of numerical concepts and to mathematical word problems requiring just such a task.

1.3.3 Conclusions about Counting

It so seems that there are two main views with regards to children's understanding of counting. On the one hand, Schaeffer, Eggleston, and Scott (1974), Baroody (1984d), Michie (1984), Fuson et al. (1983), Cowan (1987), and Saxe (1977), claim that children count not realizing what they are doing,

because they do not understand the principles of counting. They eventually understand these underlying principles by acquiring the appropriate logical mechanisms. Before then, counting is unimportant and mere parroting. On the other hand, Gelman and Gallistel (1978), Gelman and Meck (1983), Gelman, Meck, and Merkin (1986), claim that children understand counting. Furthermore, Sophian (1987) claims that children understand counting, but do not understand its relation to number as a measure.

There seems to be little doubt that pre-school children do not have a good understanding of counting. Gelman et al.'s (1978, 1983, 1986) data is very unclear and doubtful. The quotité/quantité distinction is very powerful evidence that children have to acquire basic logical skills in order to understand number and may only do so several years after beginning to count.

On balance, there is no good evidence, yet, about what leads to the acquisition of counting skills or to the understanding of the principles underlying counting. Hence, it is not clear which of the two views is right.

1.4 Addition and Subtraction

Finding out the difference between two quantities is bound to involve addition and subtraction. Addition and subtraction are basic mathematical operations that can also be used to work out the difference between two quantities. Early in life children learn to increase and decrease single quantities by adding (e.g. another brick on a pile of bricks) and subtracting. It is important to consider how it is that children relate these operations to their understanding of number so that we can subsequently determine how it is that they use addition and subtraction to compare two unequal quantities. A review of the literature on addition and subtraction will lend insight into the

extent to which children understand these operations and the age at which they begin to exhibit such an understanding.

1.4.1 Relating Addition and Subtraction to the Increasing and Decreasing of Single Quantities

From a very young age, children probably understand addition and subtraction through the experience of increasing or decreasing a given amount of concrete material. In an attempt to see whether young children understand the relationship between number and the solution to addition and subtraction problems, Starkey and Gelman (1982) asked 3, 4, and 5-year-old children to determine how many pennies were being held in the experimenter's open hand. Once this was established, the experimenter then either added or removed some pennies from the same hand holding the initial set, while specifying the number of pennies being added or subtracted. Each child was then asked how many pennies were there altogether in the hand. Throughout the various trials, the children were allowed to see only the pennies being added or removed, but not allowed to see how many remained in the experimenter's hand as these were covered.

Children in all age groups performed better with small numbers than with large numbers (i.e. 2+1 vs. 2+4), with the percentage of correct answers increasing with age (3-year-olds: 73% correct on "2+1", but only 7% correct on "2+4"; 4-year-olds: 100% correct on "2+1", but only 25% correct on "2+4"; 5-year-olds: 100% correct on "2+1", but only 69% correct on "2+4"). However, it was observed that most of the children used fingers to represent the screened pennies and yet others counted aloud, either imagining the pennies or working out the sequence of number names. They tended to count even when the pennies were out of sight. Notwithstanding, and even though the counting could be considered a confounding factor, Starkey and

Gelman concluded that children between the ages of three and five can carry out simple addition and subtraction, and understand the relationship between number and the solution to these problems. In other words, the experimenters attributed the children's success in solving these tasks to an understanding of addition and subtraction, and not to the way that children found the solution by means of counting. The children seemed to have been counting quantitatively, in which case their understanding of number as related to addition and subtraction seems to have been quite well developed. Apparently, Starkey and Gelman (1982) thought their results conclusive with regards to increasing or decreasing single quantities. One would assume the same would apply to the comparison of two unequal quantities. This will be investigated further on in this thesis.

In another study, Starkey (1983) set out to prove that even very young children, ages 24 to 35 months, were capable of exhibiting an understanding of simple addition and subtraction. Starkey asked the children to put two, three, or four objects in a container. Then, he proceeded to either add more objects to the container; take some objects out; or leave the number of objects unchanged. The child was subsequently asked to remove all the objects one by one from the container. The container was constructed in such a way that the child could only remove one object at a time. The number of reaches made by the child into the container was considered the physical manifestation of the child's mental solution to the problem. Starkey assumed that the mere fact that the child reached into the container was an indication that the child was under the impression that at least one object was there. The task did not require the child to understand the number words "one", "two", and "three".

Most of the children performed a numerically correct search on most of the problems that involved small numbers, that is, less than four. The performance was above chance level. There was also no difference between

problems involving addition and those involving subtraction. Starkey concluded that even very young children have a basic understanding of these two mathematical processes. However, it is possible that the children successfully completed the task by counting the times that the experimenter reached into the container to add or subtract an object. Objects in the first addend should be put in simultaneously in another condition in order to clear this possible confounding factor.

Hughes (1986) set out to assess children's understanding about the relationship between number and the solution to simple addition and subtraction problems, involving the increasing or decreasing of a single quantity. To measure this understanding, Hughes gave children (ages 2;9 - 4;11) 5 different tasks.

In the first task, the experimenter put some bricks into an open box, and asked the child how many bricks were in the box. After the child had responded, the experimenter added or removed some bricks, and again asked the child how many bricks were in the box. The second task resembled the first except for the fact that the box was closed so that the child could see the bricks being added or taken out but could not see how many there were altogether in the box. The third and fourth tasks were hypothetical in nature (the objects were not physically visible to the child). Hughes asked a "hypothetical box" question: *"If there was one brick in the box and I added two more, how many would there be?"*. He also asked a "hypothetical shop" question: *"If there was one child in a shop and two more went in, how many children would be in the shop now?"*. The fifth and final task presented the child with a formal mathematical problem: *"What does one and two make?"*.

Once again, children performed well with small numbers (i.e. three or less) but not with larger numbers. Hughes also found that when the numbers were small, the children, regardless of age-group, performed as well on

subtraction problems as they did on addition problems. When numbers became larger, however, addition was easier for the child to perform. Results indicated that visible-object problems were the least difficult for the children; followed by the hypothetical problems, with an intermediate level of difficulty; and lastly by the formal problems, the most difficult.

Hughes concluded that concrete material and action help children solve these kinds of problems. He argued and later postulated in 1986, that the problem of children with arithmetic must be explained in terms of abstraction, not linguistics.

Hughes' conclusions are strongly supported by the fact that when children start school at around age 5, they are able to perform simple addition and subtraction operations so long as these involve specific objects, people, and events. When children in this age group are presented with similar addition and subtraction operations without any reference to specific objects, people, and events, they are unable to perform the task successfully. Children at age 5 have a concrete understanding of things and events enabling them to perform simple addition and subtraction operations; but since they have not yet acquired abstraction, they are inhibited from performing other non-specific types of addition and subtraction operations, such as those involved in word problems. Hughes concludes that children's concepts are formed through an interaction with the physical environment. The child arbitrarily responds to a formally presented mathematical problem, and yet quite easily solves the same problem when it is "made concrete", related to specific, physical things.

Hughes, like Starkey and Gelman (1982), emphasizes the importance of fingers and argues that fingers give a concrete reference to the use of the language of arithmetic. Hence, fingers link up the abstract and the concrete since they can be both representations of objects and objects in their own right.

Hughes begins a connection between counting and addition and subtraction. There are two different ways in which the understanding of the counting system and of addition and subtraction may be connected.:

(1) Counting can act as a solution to addition and subtraction problems. The child can exhibit counting strategies to concrete or verbal problems. These strategies will be discussed in the next section (1.4.2).

(2) The additive composition of number, or realising that numbers are additive combinations of other numbers, (i.e. 7 is the additive combination of 4 & 3, and 5 & 2 [and of 3 & 4 and 2 & 5], which means that if 2 is subtracted from 7 the result is 5), is another way in which children begin to understand the connection between the counting system and addition and subtraction. In order to understand the additive composition of number, the child must grasp part-whole relations, and also the inverse relation between addition and subtraction, as well as commutativity. Part-whole relations will be discussed in subsequent chapters, as they are crucial to this thesis' central issue regarding children's quantitative comparison. Inverse relations between addition and subtraction and commutativity relations will not be discussed, as they are not relevant to the central issue of this thesis.

1.4.2 Children's Strategies for Adding and Subtracting

What children learn in school about mathematics is not the beginning but the continuation of the development of mathematical thinking. Children are continually learning from their environment. From a very young age they are already dealing with quantities: counting, dividing whole into parts, and adding or subtracting objects to make "more" or "less".

When comparing quantities, very young children exhibit certain additive strategies based on counting. In order to investigate the level of children's understanding of correspondences and numerical differences,

children's strategies for establishing correspondences between disjoint sets must be examined. Fuson (1982) observed children's behavioural patterns when performing addition and subtraction operations versus their reaction time data. Previous research had reported either the former or the latter separately, but never one aspect in contrast with the other. Based on her observations, she developed a theory claiming that children's mathematical behavioural patterns develop from a "counting-all" strategy (i.e. counting all objects of the final quantity) to a "counting-on" strategy (i.e. counting on from the initial quantity to the final quantity). This remarkable counting-on strategy is one that seems not to be used in school, but to be "invented" or used spontaneously by children for their own purposes and improves with age.

Part of the data supporting this idea came from studies conducted by Groen and Resnick (1977), who observed the development of children's use of the counting-on strategy. In their experiment 5-year-old children were initially taught to solve simple addition problems, presented in written form, by "counting-all" with blocks. For example, they were told to respond to "2+3" by counting out two blocks, then by counting out three blocks, and then by counting the combined set. After several practice sessions spread over a period of a few weeks, the children were presented with more addition problems and told to solve them without the blocks. Groen and Resnick observed, via latency measures, that half of the children made a clear and spontaneous transition from mentally counting-all the items to counting-on-from-the-larger-addend (e.g. responding to "2+3" by counting on from 3). This is known as the MIN Strategy (Groen and Parkman,1972), where MIN stands for minimum, which is exactly what the child wishes to count (i.e. the smaller of the two addends). The experimenters concluded that at age 5, children will solve most addition problems through a MIN strategy, which will develop spontaneously, without the aid of formal instruction. For the

purposes of this investigation, the MIN strategy will only be dealt with in the light of counting-on as a measuring strategy.[1]

As soon as a child begins to count-on, there is a recognition of the fact that the addend consists of a quantity. Using the counting-all strategy, does not necessarily demonstrate a recognition of the addend as a quantity. Rather, these strategies suggest that a mathematical development takes place when this occurs and is very relevant towards investigating how children work out the difference between two quantities. For example, one strategy the child may use for working out the difference between, for example, A and B, where B is smaller than A, is to add on to B to make it the same as A. By noticing how much should be added on, the child solves the difference between the two quantities. This is also particularly relevant to Equalize-type problems. In order to use the latter strategy, the child must know the strategy of counting-on.

Baroody (1984c) investigated children's solving of addition problems and their transition from a concrete counting strategy to a mental counting strategy. He presented 17 children ranging in age from 4;11 to 6;7 (median age 5;4) with three types of tasks.

The first task, or the initial addition task, made use of several 5 x 8 inch cards each of which had one unsolved addition problem typed on it in horizontal form. The addition problem on the card was read to the child. For example, the problem "5 + 1" would be read as it was written and also as follows: "This says five <u>and</u> one. How much are five and one altogether?". The child had the option of either solving the problem mentally or using concrete aids (blocks or fingers). If the child could not solve the problem concretely, the child would then be taught concrete counting-all and would then be asked to imitate that particular strategy. <u>Concrete counting-all</u> refers

[1]Apart from Groen and Resnick (1977), already mentioned, there is extensive literature on the MIN method using latency measures, (particularly that of Hitch et al.,1987), which will not be evaluated, as it is not directly related to the questions asked in this thesis.

to children's use of concrete countable objects such as fingers or blocks, which are counted one by one to represent an addend. This same process is repeated a second time for the second addend. Then all the fingers or countable objects put out are counted to determine the sum. Children were deemed to perform successfully if they obtained the correct answer with or without objects. Children would be given half credit if they knew a strategy but simply miscalculated. Children were deemed to have responded incorrectly if they had to be taught the concrete counting-all strategy. Success was defined as 5/6 trials correct.

The second task was the addition-practice task, which consisted of ten addition trials presented randomly. All these ten trials had the smaller term first. This was done in order to differentiate between mental counting strategies. The child was encouraged to use mental strategies, even guessing.

Mental counting strategies require a single keeping-track process, as Baroody explains. Mentally counting-all-starting-with-the-first-addend entails starting with "one", counting up to the cardinal value of the first addend, and then continuing the count for a number of steps to the cardinal value of the second addend (e.g., 2 + 4: "1,2; 3(1 finger up), 4(2 fingers up), 5(3 fingers up), 6(4 fingers up) = 6"). Mentally counting-on-from-the-first-addend is a short cut to counting-all-starting-with-the-first-addend, as it starts with the cardinal value of the first addend (e.g., 2 + 4: "2; 3(+1), 4(+2), 5(+3), 6(+4) = 6"). With the strategy, mentally counting-all-starting-with-the-larger-addend, a child starts with "one", counting up to the cardinal value of the larger addend and then counting on from there while the smaller term is enumerated (e.g., 2 + 4: "1,2,3,4; 5(+1), 6(+2) = 6"). This strategy minimizes the cognitively demanding double count required by mental addition strategies. Mentally counting-on-from-the-larger-addend, previously referred to as the MIN strategy, is a shortcut to this procedure and it reduces mental effort, as it

starts with the cardinal value of the larger term and hence it is the most economical mental counting strategy (e.g., 2 + 4: "4; 5(+1), 6(+2) = 6").

Except in the case of a minor counting error, children were then encouraged to solve the problem via blocks or fingers. Only a child's predominant strategy was reported. The "predominant strategy" was defined as a strategy that a child used more often than other strategies and used it on at least 3/7 of the trials.

The third task was the commutativity task. As this task is not relevant to the theme of this thesis, it has been omitted from the general review of the literature.

Baroody carried out this experiment as a longitudinal study. He found that in the initial addition task, 14 out of 17 children had to be shown the concrete counting-all strategy. Two of the three, who did not have to be shown the concrete counting-all strategy, used mental strategies instead.

After a child learned the concrete counting-all strategy, the child then adopted shortcuts of concrete counting-all or mentally counting strategies. After this, the child did not revert to using a less advanced strategy. Baroody found, however, that the transition from concrete to mental addition was not an easy one and was often delayed.

Only five of the children made the transition to relying on mental strategies; one child had already made the transition. Only three children who adopted mental strategies as their dominant approach did not do so until after the fourth session.

Eleven out of 15 children who initially relied on concrete counting-all strategies never adopted a concrete counting shortcut. One out of 15 children used a concrete counting shortcut as her predominant strategy for a single session but then fell back on using concrete counting-all before inventing a mental strategy. Three out of 15 invented mental strategies without first relying on concrete counting shortcuts.

When children mentally computed sums, their preference was for strategies that disregarded addend order. Among the mental strategies, counting-all-starting-with-the-larger-addend was used more frequently than mental counting-all-starting-with-the-first-addend, among the mental strategies. For children who invent counting-all-starting-with-the-larger-addend, counting-on-from-the-first-addend makes little sense as the next developmental step because it does not minimize the number of steps in the cognitively demanding keeping-track process. This suggests that even among the kindergarten age children just developing a mental addition strategy, there is a tendency to minimize the cognitively demanding keeping-track process by starting with the larger addend. Hence, for some children, counting-all-starting-with-the-larger-addend may be an important transitional step.

It is interesting to note that the strategy of counting-on-from-the-first-addend was rarely observed. Baroody notes that only two children used this strategy. Baroody points out that for 4 + 5, a five step keeping-track process is relatively easy because the child only needs to count until all five fingers of a hand are extended.

The strategy of counting-on-from-the-larger-term was not always easily adopted (Baroody, 1984c; Carpenter and Moser, 1984). Two of the children abandoned the use of all other strategies, when they began using the strategy of counting-on-from-the-larger-term. However, three of the children used a mixture of strategies before relying completely on this latter one.

Baroody's study resulted in different conclusions from those of Fuson (1982) (see Section 1.3). Fuson had argued that the method of keeping-track was an important transitional step to those methods that involve matching the count or making a double count. However, Baroody almost never observed counting strategies that involved creating a model to facilitate the keeping-track process. However, he agreed with Fuson that perhaps there was an

intermediate form between a true counting-entities strategy and a true mental strategy, since even though fingers are extended successively, the child may not actually count the fingers as they are extended, but may use an implicit finger pattern to determine when to stop the sum count.

Baroody concluded from his results that one cannot assume that a child will take for granted the concrete strategy for computing sums or overestimate the extent of such knowledge among children entering school, even though his study indicates that many children just beginning school do have informal addition strengths. Very few children, in Baroody's experiment used a concrete counting-all strategy to calculate sums. The majority of the children on Baroody's task needed repeated demonstrations, (at least two or three) of the concrete counting-all strategy before they mastered it.

Baroody (1984c), and Baroody and Ginsburg (1986), propose that one reason why concrete counting-all is so difficult for some young children to learn may be that such a strategy more directly models a union-of-two-sets view of addition - a binary conception. Children's view of addition may be that of a change of state - a unary conception. Some children define addition in terms of "action schemes" or changes of state (Baroody and Ginsburg,1983); this is also considered as treating addition as a unary operation (Weaver,1982). Hence 3 + 2 would be interpreted as "three and two more", which is a unary conception, rather than as combining the cardinal number three and cardinal number two, which is a binary conception. Some children may have treated the larger addition trials as a counting task, as they may have interpreted addition as a procedure for counting across two sets of objects or numbers rather than a procedure for combining quantities. Baroody proposes that, if this is the case, then a good way to teach addition to children, especially to low-functioning children, would be to introduce it in a manner consistent with their unary conception - for example, by starting with problems in which 1 is added to an initial set.

Baroody's investigation was a successful one. It provided a control question, which was the initial addition task, and an addition-practice task, which brought out a whole investigation of children's additive strategies. Furthermore, his conclusions lend insight into the child's cognitive processes when performing addition problems.

Baroody (1984a) also investigated children's solving of subtraction problems. He explored the counting-down procedure, parallel to the counting-on procedure for addition. Counting-down involves (1) stating the larger number, (2) counting backward a number of times equal to the smaller numbers, and (3) announcing as the answer the last number counted. This obvious demand for two simultaneous processes taking place at the same time explains the complexity of subtraction vis-à-vis addition. An example of this is illustrated when a child is asked to solve 5-2. Starting with "5, 4 (that's one taken away), 3 (that's two taken away), = the answer is 3". As one can see, counting-down involves a mental number-before-N procedure. However, with N-1, the child just has to know what number came before another in the number word sequence. With N-2, N-3, and so on, the child must be able to count backward from a specified point within the number word sequence. If a child cannot use "preceding number" and "next number" relationships, then the child cannot compute N-1. This will result in difficulty in employing the counting-down strategy, as the N-1 algorithm is an essential component of counting-down. If a child cannot compute the N-1 algorithm, then this will be virtually impossible when confronted with minuends equal to or greater than 2. Moreover, if a child cannot count backwards, then the N-1 procedure cannot be extended into a counting-down procedure. It is obvious that in order to count down, the child must know how to count backwards and do so with ease. Baroody (1983) pointed out that if the child is not efficient at counting backwards, then counting-down and counting backwards simultaneously may be too great a load. If a child cannot consciously keep

track of the number of times backward counting is used, then it is not possible for the child to perform the strategy of counting-down. To make this cognitive process even more difficult and complicated, the child must be able to count backward a specified number of steps. The procedure (5-2: five, four is one, three is two), entails a forward count to keep track of the smaller numbers or some other keeping-track method often dealing with the backward count by using fingers, which in effect involves a forward count (i.e. 5-2: five, four (first finger up), three (second finger up) (Fuson,1982). Counting-down thus involves two simultaneous processes that in effect go in opposite directions.

Baroody (1984b) explains the problem in terms of the size of the smaller number in the subtraction operation. For example, when a child is given the problem 9-2, the double count or any other simultaneous keeping-track process is relatively manageable, as it involves only two steps. However, when the child is confronted with a more difficult operation, such as 9-7, the keeping-track process becomes relatively unmanageable, as it involves seven steps. For a problem such as 19-17, the process becomes virtually impossible. It has also been suggested that the difficulty may not only involve the smaller number in the operation, but also the larger number in the operation. As Baroody has pointed out, counting backward from 20 is sometimes more difficult for primary school children than counting backward from 10.

It is important to note that subtraction's counting-down method is more difficult than addition's counting-all or counting-on methods. Even though these two types of addition methods do involve simultaneous processes, at least these two simultaneous processes go in the same forward direction. Subtraction's simultaneous processes, as we have seen, go in opposite directions. Counting backward is more difficult than counting forward.

Children tend to substitute the counting-down procedure with a counting-up procedure, as a result of the difficulty they encounter with the

former procedure when confronted with subtraction problems involving larger numbers.

Woods, Resnick, and Groen (1975), for example, claimed that children first discover the counting-down procedure and later discover the counting-up procedure. Counting-up involves starting with the smaller number and counting forward until the larger number is reached, while keeping-track of the number of steps in the forward count. For example, 19-17 would be calculated in the following manner: 17; 18 (is one), 19 (is two) = so that answer is 2. At times this counting-up procedure is more economical than counting down. For example, when the smaller number and the larger number are relatively close, as in for example, 9-7, counting-up greatly reduces the amount of keeping-track (e.g. by double counting) that is necessary, as it involves only 2 steps versus 7 steps. When the smaller number and the larger number are relatively far apart, as in for example, 9-2, counting-down greatly reduces the amount of keeping-track that is necessary, as the keeping-track process involves only 2 steps versus 7 steps in the case of counting-up. Woods et al. (1975) report that children by the age of 7-8 can select the more economical procedure for a given subtraction problem.

Neither of the two procedures is more basic than the other. However, the schools do emphasize counting-up more than counting-down. For example, the Wynroth approach suggests that subtraction not be referred as "take-away", and involves teaching the child to count up. This approach is also known as the missing-addend approach. Baroody reports that even first graders employ the counting-up procedure to solve subtraction problems. On the other hand, many children do continue to use the counting-down procedure to solve subtraction problems, despite being taught the counting-up procedure at school. Children tend to continue using informal methods that make sense to them, rather than adopting those methods taught in school, which they find difficult to understand conceptually. They prefer

counting-down as it is similar to their informal concept of take-away and represents a natural extension of their number-before-N procedure. Counting-down is also encouraged by the types of subtraction problem which young children are given at school, usually involving small subtrahends, so that counting-down is the more economical procedure. Hence teachers can expect children to use a counting-down procedure when they do mental subtraction.

Another strategy used by children in subtraction is the separating-from strategy. This strategy consists of (1) representing the larger number, (2) removing a number of items equal to the smaller numbers, and (3) counting the remaining items to determine the answer. An example of this is illustrated when a child is asked to solve 5-2. Firstly, the child counts out five fingers or objects (making five marks), secondly, the child counts and removes two of the items (crossing out two of the marks), and thirdly, the remaining items are counted (marks) = "Three".

Siegler (1987) also investigated children's subtraction strategies and videotaped the children's subtraction performances. He tested 17 five-year olds and 17 six-year-olds. Children were presented with 25 subtraction problems in which the smaller number (the subtrahend) ranged from 1 to 5 inclusive. The difference also ranged from 1 to 5. The 25 problems were presented during a three day period; the child being given about eight or nine problems a day. Afterwards the child was presented with the same problems again, but in a different order.

Siegler observed that the strategy, counting-down-from-the-larger-number-the-number-of-times-indicated-by-the-smaller-number, was among the most frequent strategies used by children in subtraction problems. Counting-down, however, was not observed to be very frequent in Siegler's experiment. Only 38% of the correct answers to the 25 problems was accounted for by the total number of counts made in a problem (counting-up-

to-the-larger-number and/or counting-up-to-the-difference-between-the-two-numbers and/or counting-down). The size of the smaller number accounted for 57% of the total number of counts. Siegler argues that a reason why counting-down was not influential, may be due to the extremely high error rate using this procedure. More than 80% of the children's errors in counting on the counting fingers trials came when they were counting-down rather than up, even though counting-down occurred in less than 30% of the total counts. Siegler suggests that the higher error rate could be attributed to both the unfamiliarity of counting-down and to the short term memory demands imposed by the counting-down process. As previously reviewed, (Baroody, 1984a), counting-down needs keeping-track of the counts and keeping-track of the intended arrival point.

Counting-down-the-number-of-times-indicated-by-the-smaller-number does become difficult when the subtrahend increases in size. However, in Siegler's task, this was not the case when the subtrahend equalled five. When this was the case, children instead of decreasing the number one at a time, lowered the whole hand (i.e. all of the five fingers). When the children did this, they answered accurately and rapidly.

Siegler also reported predictor variables that accounted for 84% of a child answering the 25 subtraction problems correctly:

(1) Size of the smaller number (the subtrahend)

This predictor variable accounted for 7% of the 84% correct answers given by the children on the 25 subtraction problems.

(2) Smaller number equal to five

This predictor variable accounted for 13% of the 84% correct answers given by the children on the 25 subtraction problems.

(3) Associative strength of the correct answer to the exact inverse addition problem

An example of this predictor variable, is that for the problem 5-4, the associative strength linking the problem to the exact inverse addition problem is 1+4=5. This predictor variable accounted for 64% of the 84% correct answers given by the children on the 25 subtraction problems.

There are other predictor variables which Siegler did not report, but are worth mentioning.:

(4) Size of the larger number (the minuend)

(5) Size of the sum of the larger (the minuend) and smaller numbers (the subtrahend)

(6) Difference between the larger (the minuend) and smaller numbers (the subtrahend)

Siegler does not account for the possibility of children making number-fact responses. In his task, it was observed that when the subtrahend equalled five, children lowered the whole hand, and calculated the answer accurately and rapidly. It must be taken into account that Siegler does not give any leeway to attribute this type of observation to number-fact.

1.4.3 An Addition and Subtraction Model

Siegler and Shrager's (1984) model of distribution of associations outlines children's performance of addition and subtraction. The model is so called because within it, errors, solution times, and overt strategy use are all functions of a single variable: the distribution of associations between problems and potential answers. The process involved in this distribution of associations model can be represented as consisting of three phases: retrieval, elaboration of the representation, and counting. Development in this distribution of associations model is influenced by three factors: pre-existing associations from the counting string, frequency of exposure to the problems, and the sum of the two addends.

Siegler and Shrager presented 50 trials of 25 subtraction problems to 34 children of ages 5-to-6-years-old. The problems were the direct inverse of the addition problems. So for example, a problem given for addition in the form of A+B=C, would then be given for subtraction in the form of C-B=A. The children were videotaped.

Results indicated that children used visible and audible (i.e. overt) strategies for subtraction in the same way as they did for addition problems. Such strategies observed were:

(1) Counting fingers

This strategy consisted of raising the fingers to represent the larger number (the minuend), as well as lowering the fingers until the number of fingers being tucked away equalized the smaller number (the subtrahend), and lastly, counting the numbers making up the difference. Both the raising and lowering of the fingers to count was done either all at once or one at a time. However, when the lowering took place, it was usually carried out one at a time.

(2) Counting

This strategy consisted of more or less the same as the counting fingers strategy with the exception that the children did not use fingers or other objects as aids to counting. They just counted aloud.

(3) Fingers

This strategy consisted of raising the fingers to represent the larger number (the minuend) as well as lowering the fingers to represent the smaller number (the subtrahend), and finally giving the answer without counting the fingers left as the number making up the difference.

(4) Retrieval

This strategy involved no counting aloud or use of fingers or other objects as aids to counting. Children simply stated the answer.

Hence, children subtracted more frequently on trials where they used the retrieval strategy. This strategy was used on 58% of the trials. Children subtracted more accurately on trials where they used the fingers strategy. The counting fingers strategy was the next most accurate strategy, and retrieval and counting were the least accurate. Another interesting result concerned the mean solution time. Retrieval was a faster strategy than the fingers strategy, which in turn was faster than the counting strategy, which in turn was faster than the counting fingers strategy. As in addition, there was a relationship between overt strategy use and errors: percentage of overt strategy use was a reliable predictor of percentage of errors. Another interesting result concerned the mean solution time. As in addition, the longer the mean solution time on a problem, the higher the percentage of overt strategy use on that problem.

Siegler and Shrager (1984) concluded, on the basis of their results, that in subtraction, as in addition, the types of strategies that children used, the relative solution times for the strategies, the correlations between percentage of errors, mean solution times and percentage of overt strategy use on each problem, as well as the errors and solution times on retrieval trials, all matched the pattern predicted by the distribution of associations model. Errors and solution times on the trials, like overt strategy use, could be derived from the distribution of associations model. However, errors and solution times on counting and counting-fingers trials would be expected to relate to the size of the numbers being counted. In addition, the size of the numbers being counted corresponds to the sum, whereas in subtraction, it corresponds to the size of the larger number. Siegler and Shrager concluded, that the frequency of overt strategy use was likely to be an accurate predictor of errors and solution times on counting and counting-fingers trials, especially when the contribution of the larger number was partialed out.

Siegler and Shrager's model of distribution of associations of children's performance on addition and subtraction has a basic underlying limitation. It does not account for other additive and subtractive strategies (previously discussed in Section 1.4.2), other than the counting-fingers strategy, the counting strategy, the fingers strategy, and the retrieval strategy. It is not clear whether they consider their mentioned strategies exhaustive.

1.4.4 Conclusions about Addition and Subtraction

The link between counting and addition and subtraction in order to compare two unequal quantities relies heavily on the fact that children have to learn to choose particular types of counting appropriate to the solution of particular types of problem. For example, a child may choose to count backward to solve a subtraction problem, or to count forward to solve an addition problem. Hence, counting algorithms are used to solve addition and subtraction operations. According to Baroody (1984d), Ginsburg (1982), Groen and Resnick (1977), and Ilg and Ames (1951), preschool age children have already discovered that the counting sequence helps them to solve simple mathematical problems.

Baroody (1984a; 1984b; 1984c), like Ginsburg (1982), and others discuss ways in which children, even when they are presented with basic addition and subtraction combinations, still tend to use their informal counting strategies to obtain the answers. Baroody (1984a; 1984b; 1984c), like Groen and Resnick (1977), further explains that as time goes on, children invent sophisticated counting strategies for adding and subtracting. The teaching of formal mathematics gives the child a developmental basis for more advanced strategies. For example, a child may deduce from 4+4=8, that 5+4=9, because 5 is 1 more than 4, or s/he may also deduce, for example, that 6-4=2 because 2 plus 4 is 6. Hence, a child may be able to deduce numerical

relationships from number-fact, which is knowledge about other numerical relationships, or from just a general knowledge of number. As mentioned previously, one reason why primary school age children initially rely on their informal mathematical procedures may be that these are meaningful to them.

1.5 General Conclusions

Having established children's understanding of quantity in general within the context of one-to-one correspondence, counting, and addition and subtraction, the next sections (1.6 - 1.8) will deal with children's understanding of quantitative comparison within the context of word problems. One-to-one correspondence, counting, and addition and subtraction are ways that children use to compare quantities. The next sections (1.6 - 1.8) will investigate the relevance of these afore-mentioned ways to word problems.

1.6 Word Problems

Another way of investigating this thesis' central issue of how children compare quantities and how they work out the difference between them is through word problems. Word problems are sums embedded in stories. They have been extensively used to test the mathematical abilities of children. There are four categories of word problem: Compare, Equalize, Change, and Combine. Two of these are essential for understanding the difficulty that children have in comparing two unequal quantities and how they work out the difference between them. These are Compare word problems and Equalize word problems. The other two types of word problems are not related to the theme of this investigation, as they do not test children's understanding of quantitative comparison, but they nevertheless will be briefly

mentioned and described. These are Change word problems and Combine word problems.

Work, which has been reviewed already (Section 1.3), suggests that the difficulties children have in comparing two unequal quantities might be caused by a failure to count. The crucial question then is whether children depend on counting in these word problems and whether any difficulty they might have with them is specific to counting and number. Furthermore, following from this crucial question is a second crucial one, which is, whether these difficulties come up as much with continuous, as with discontinuous material.

Compare and Equalize word problems deal directly with how children compare two unequal quantities. These two types of problem are entirely equivalent in nature, requiring the same computation and calculation, involving the same mathematical operation in working out the difference between two quantities, and yet children seem to have difficulty with one and not so much with the other.

Examples of Compare problems are the following:

a. **Mary has 6 books. John has 4 books.**
How many more books than John does Mary have?

b. **John has 5 comics. Mary has 2 comics more than John.**
How many comics does Mary have?

In this category two distinct quantities are compared to find the difference between them (i.e. 6-4=X). Hence, a Compare problem deals with a "between-type" relation, as a comparison is made "between" two distinct quantities. Such problem also involves a "static" relationship between two quantities. Static relations embody conceptual knowledge about comparisons

involving sets of objects and do not pose a transformation between the quantities.

An example of an Equalize problem is the following:

Mary has 6 marbles. John has 4 marbles.
How many more marbles does John need,
to have as many marbles as Mary?

Two given sets are compared, and solving the problem involves finding the amount by which one of the sets must be changed to make it equal to the other (i.e. 4+X=6). Hence, an Equalize problem deals with a "within-type" relation, as a comparison is made of the same quantity "in" two different states. Such problem also involves an "action" relationship, where one set is joined to or separated from another set. Action relations are different from static relations in that they embody actions that cause increases or decreases in some quantity and pose a transformation within the quantities.

From a rather different point of view, an example of a Change problem is the following:

John had 4 marbles. He got 2 more.
How many marbles does John have now?

Problems in this category have an initial quantity and some direct or implied action causes a change in the quantity. They involve an "action" relationship where one set is joined to or separated from another set. The direction of change can either be an increase (a join problem), or a decrease (a separate problem).

An example of a Combine problem is the following:

Mary has 2 apples. John has 4 apples.
How many apples do they have altogether?

Problems in this category involve two distinct quantities that are parts of a whole and therefore must be considered in combination. Hence, a "static" relationship between the two quantities is involved.

1.6.1 Evidence for the Difficulty of Compare Problems

Of the two kinds of word problem which deal with differences between quantities, there is much more evidence on children's performance with Compare problems. Equalize problems have been relatively neglected, and as will be seen, there is a lack of proper systematic comparisons between Compare and Equalize problems.

The most interesting finding resulting from research on Compare problems is the fact that Compare problems are very difficult for young children.

Riley (1981), for example, gave children sets of word problems to be solved using blocks. The basic procedure involved having children individually solve selected word problems that were read to them by the experimenter. Memorial and computational difficulties were kept at a minimum by reading the problems slowly, repeating them if necessary, and by restricting the size of the numbers in the problems such that sums were less than 10. In addition, concrete objects (blocks) were provided for children to use in solving the problems.

Results demonstrated that for the 4- to 5-year-old group, Compare problems (.11=[Proportion of Children who performed correctly using

Objects]) were more difficult to solve than either Combine (.61=[Proportion of Children who performed correctly using Objects]) or Change (.62=[Proportion of Children who performed correctly using Objects]). For the 5- to 6-year-old group, Compare problems (.19=[Proportion of Children who performed correctly using Objects]) were again more difficult to solve than either Change (.67=[Proportion of Children who performed correctly using Objects]) or Combine (.70=[Proportion of Children who performed correctly using Objects]). For the 6- to 7-year-old group, Compare problems (.72=[Proportion of Children who performed correctly using Objects]) were consistently more difficult to solve than Combine (.85=[Proportion of Children who performed correctly using Objects]) or Change (.92=[Proportion of Children who performed correctly using Objects]). Finally, children of 7- to 8-years-old performed relatively well on the Compare problems (.91=[Proportion of Children who performed correctly using Objects]), as well as on the Change (.96=[Proportion of Children who performed correctly using Objects]) and the Combine (1.00=[Proportion of Children who performed correctly using Objects]) problems.

Hence, Riley's data show that young children do find particular difficulty solving Compare problems. The evident difficulty of Compare problems for young children cannot be due just to the difficulty of the sums involved because they can often manage the equivalent sums in Combine and Change problems. Therefore, it seems likely that their difficulty lies in doing these sums in order to work out the difference between two quantities. In that case, one needs to know if they have as much difficulty in the other kind of word problem in which they have to work out the difference between two quantities - the Equalize problems.

Unfortunately, there is very little evidence on this crucial question. Carpenter, Hiebert, and Moser do indeed compare the two kinds of problem, but the Compare and Equalize problems were not equivalent. They gave 43

children, with an average age of 5 years, two kinds of Compare problem and two kinds of Equalize problem. Examples of the two Compare problems were:

(1). Ralph has *a* pieces of gum. Jeff has *b* more pieces than Ralph. How many pieces of gum does Jeff have?

(2). Mark won *a* prizes at the fair. His sister Connie won *c* prizes. How many more prizes did Connie win than Mark?

Examples of the two Equalize problems were:

(1). Joan picked *a* flowers. Bill picked *c* flowers (a<c). What could Joan do so she could have as many flowers as Bill? (Suggest, if necessary, that she pick some more.) How many more would she need to pick?

(2). Fred has *a* marbles. Betty has *c* marbles (a>c). What could Fred do so he would have as many marbles as Betty? (Suggest, if necessary, giving some away.) How many would he need to get rid of?

They found that the children did quite well on both Equalize problems and as well on Compare problem (2). However, they did very poorly on Compare problem (1).

There are two difficulties here. One is that there is no Equalize problem equivalent to the difficult Compare problem. In the difficult Compare problem the child is given the difference between the two quantities as well as one of the quantities and has to work out the other quantity. All Equalize problems involve working out the difference between two quantities.

The second problem is that Carpenter, Hiebert, and Moser's results with Compare problem (2), which was easy for the young children, seem to be quite out of line with those of other experimenters, who, as previously

shown, found that this was a difficult sort of problem for young children. So, the Carpenter, Hiebert, and Moser finding, that there was no difference in difficulty between Compare Problem (2) and the equivalent Equalize problem, does not provide a satisfactory answer to our question whether Equalize problems cause children the same difficulty as Compare problems.

Another study which involved Compare problems was done by DeCorte and Verschaffel (1987a). They observed children's errors by giving 30 children of ages 5-to-6 eight word problems and a series of Piagetian tasks, memory tasks, and counting tasks. The children were given these batteries of tests three times during the academic year. The word problems were read aloud and the children were asked: (1) to retell the problem, (2) to solve it, (3) to explain and justify their solution methods, (4) to build a material representation of the story with puppets and blocks, and (5) to write a matching number sentence. At the end of the task, the child was then asked to construct a word problem out of an unsolvable word problem.

They gave the 30 5-year-old children Compare problems in two forms:

Problem A: Pete has 3 apples; Ann has 6 more apples than Pete; how many apples does Ann have?, and

Problem B: Pete has 3 apples; Ann has some more apples than Pete; Ann has 8 apples; how many more apples does Ann have than Pete has?,

DeCorte and Verschaffel obtained the following results.: Four of the 30 children solved Problem A correctly; six of the 30 children solved Problem B correctly; about half of the children gave the answer "6" to Problem A and about half of the children gave the answer "8" to Problem B.

Thus, these experimenters confirmed the extreme difficulty young children have with Compare problems.

1.7 Theories about the Difficulty of Compare Problems

The answer to these questions might help in developing a theory of the difficulty of Compare problems.

(1) As Equalize problems are equivalent to Compare problems, in the sense that they require the same mathematical operation, do children then encounter similar difficulty with this other type problem.(?)

(2) As the difficulty children have with Compare problems has been proposed to be possibly due to children's difficulty with counting, does this difficulty then apply to continuous, as well as discontinuous material.(?)

There are three main kinds of explanation of the difficulty that children have when solving Compare-type problems, and all of them make definite predictions about the relative ease they have when solving Equalize-type problems. One hypothesis deals with the structural aspect of Compare problems and involves the difference between static versus action relationships (Riley, Greeno, and Heller,1983). A Compare problem, as previously mentioned, involves a "static" relation, as no transformation occurs among its quantities, whereas an Equalize problem involves an "action" relation, as a transformation occurs among its quantities by increasing or decreasing one of the quantities to make it equal to the other. The former-type relation is more difficult than the latter, as children are more readily capable of understanding relationships involving action than they are of those involving static relationships. This hypothesis predicts that Equalize problems should be easier than Compare problems.

A second hypothesis deals with the linguistic aspect of word problems and involves children's failure to understand the comparative word "more" (and by the same token the word "less" as well). Hudson (1983), for empirical reasons which will be presented later, has argued that children's difficulties with Compare problems stem from problems in understanding the

word "more". Since both Compare and Equalize problems involve comparative terms, this hypothesis predicts that there should be no difference in the level of difficulty of the two kinds of problem.

A third hypothesis, proposed by this author, argues that the difficulty of Compare problems can only be explained by a combination of structural and linguistic factors. According to this hypothesis a structural explanation on its own cannot account for the difficulty of Compare problems because it is possible to make the Compare problem a great deal easier without changing its essential structure by not using comparative terms, as Hudson (1983), in work to be described later, has demonstrated. However, the third hypothesis argues that there are two senses in which the word "more" (or "less") can be used. One, the sequential sense, is to compare the same quantity in two successive states: sequential comparisons involve transformations. Comparing the height of a tower of bricks, before and after some bricks have been added to it, is one possible example: comparing the level of liquid in a glass, before and after someone has drunk from it is another. The other sense in which comparative terms are used is simultaneous: this means the comparisons of two different quantities without a transformation to either. Comparing the height of two different towers of bricks or the level of liquid in two different glasses of water are examples of simultaneous comparisons.

There is some evidence from the work of Moore and Frye (1986), which is to be reviewed later, that children understand comparative terms used in the sequential sense more easily than in the simultaneous sense. They understand the word "more" when it refers to an addition to one quantity better than when it refers to the difference between two quantities.

The reason why this difference in understanding the two senses of comparative terms is relevant to Equalize and Compare problems is that the wording in Equalize problems is in the sequential sense while in the Compare problems it is in the simultaneous sense. In Equalize problems, the child is

asked to imagine a transformation to a single quantity (*"How many more marbles does John need, to have as many marbles as Mary?"*), while in the Compare problems he is asked to estimate the difference between two unchanged quantities (*"How many more marbles than John does Mary have?"*).

Therefore, this third hypothesis also predicts that Equalize problems should be a great deal easier than Compare problems.

1.7.1 Structural Hypotheses about the Difficulty of Compare Problems

There have been several attempts to account for the difficulty of Compare problems in terms of the structural characteristics of these problems. There are three structural hypotheses about the difficulty of Compare problems. One is the hypothesis of Riley, Greeno, and Heller (1983), the second is the hypothesis of Briars and Larkin (1984), and the third is the hypothesis of Silver and Thompson (1984).

1.7.1.1 The Riley, Greeno, and Heller Model

The Riley, Greeno, and Heller model is concerned with factors that account for children's difficulties in word problem solving. These factors include the complexity of conceptual knowledge about the problem domain and the sophistication of problem-solving procedures. These factors are important to understanding how children compare unequal quantities and how they solve the difference between them.

Riley et al. base their hypothesis on an extensive review of the research on addition and subtraction word problems and factors that influence their difficulty, such as the works of Carpenter, Hiebert, and Moser

(1981), Fuson (1979), Nesher and Teubal (1975), Vergnaud (1982), Hudson (1983), and others, previously described.

Unfortunately, Riley, Greeno, and Heller take as one of their starting points, the assumption that Equalize problems are easier than Compare problems. They base this assumption entirely on the Carpenter, Hiebert, and Moser study, and yet as previously explained, there is no basis from this study for concluding that the Compare problems are the harder of the two. There was no Equalize problem equivalent to the difficult Compare problem in this study.

Riley et al. begin by defining a word problem as involving identification of some quantities and description of a relationship among them. There are certain categories of word problems. These categories, previously defined, are Change, Equalizing, Combine, and Compare. The global difference between all these categories involves the "semantic structure", according to Riley et al. Semantic structure, as they interpret, means conceptual knowledge about increases, decreases, combinations, and comparisons involving sets of objects. Change and Equalize word problems describe addition and subtraction as actions that cause increases or decreases in some quantity. Combine and Compare involve static relations between quantities.

Riley et al. distinguish between problem schemata, action schemata, and strategic knowledge. Problem schemata consist of elements and relations among those elements. Action schemata refer to planning operations to solve the problems. They present four types of action schemata: the make-set, the put-in set (virtually makes more), the take-out set (virtually makes less), and the count-all set. Strategic knowledge refers to representation of production rules organized in a way that permits top-down planning (choosing a general approach to a problem; then deciding about action that is more specific; and only then working out the details). These

three components of knowledge are needed for successful performance in the domain of word problems. If children find it difficult to solve word problems, it could be due to a failure in any one of these components of knowledge.

From their review, Riley et al. developed three types of model that simulate the different levels of children's performance on word problems. These models vary in the way that information is represented and the way in which quantitative information is manipulated. Model One (1) understands quantitative relations by means of a simple schema that limits its representations of problems to external displays of objects. Model Two (2) has a schema for maintaining an internal representation of increases and decreases in the sets of objects manipulated. Hence, the process of building this representation is relatively bottom-up in the sense that it depends on the external display of objects. Model Three (3) also has a schema for representing features internally, but can use its schema in a more top-down way than Model Two (2). A direct understanding can be achieved independently of the external display of objects. Models Two (2) and Three (3) have a richer set of action schemata for producing and manipulating quantitative information and a richer understanding of certain relations between numbers than Model One (1). It is Model Three (3), however, that enables an understanding of part-whole relations. These models provide hypotheses about how changes occur, developmentally, in children's ability to understand relationships among quantities and how to use representations of these relationships to solve problems. Riley et al. believe that successful problem-solving performance by children and adults depends on their understanding of certain concepts, which are characterized as schemata used in representing information in problem situations.

Riley et al. described three main ways in which conceptual knowledge and procedural knowledge interact when a child compares two unequal

quantities and works out the difference between them. One is through the role of schemata in the selection of actions. In the models, both problem schemata and action schemata are necessary for relating the problem statement to the actions required to solve the problem. Problem schemata are involved in interpreting the problem text. They range from a simple schema used to represent quantitative relations to a more complex Change schema, and finally to the schemata required for representing the complex relationships in Combine and complex forms of Compare problems. In all these models, the schemata is associated with goals either to change the current problem situation, to obtain some information from the problem situation and hence planning procedures, or to identify an action whose consequence matches the current goal. At times one of three things may happen. There may be a direct match between the goal and one of the model's action schemata while at other times, additional schemata are required to infer important relations in the problem situation before the appropriate action can be chosen. Secondly, there may be an interaction of conceptual and procedural knowledge involving the use of schemata to monitor the effects of selected actions on a problem situation. Lastly, there may be the influence of conceptual knowledge on the actions that get selected.

Solving a word problem, note Riley et al., requires an interaction of conceptual and procedural knowledge. Once a model has represented a problem situation (problem schemata), it must have some way of relating this representation to its problem-solving procedures (action schemata). Furthermore, in addition to problem schemata and action schemata, the models also have a strategic knowledge for planning solutions to problems. Riley et al. also suggest that a child's conceptual knowledge of the relations between quantities in a word problem is related to the acquisition of more

efficient counting procedures. Hence, necessity and efficiency are the two motivations for acquiring a more advanced and sophisticated schemata.

The different types of word problems can be categorized into these models. These models predict a difference between Equalize and Compare word problems. Equalize word problems are represented in Model Two (2)'s schema for representing quantity, as a change is represented by an increase or a decrease in the amount manipulated. Compare word problems, on the other hand, are represented in Model Three (3)'s schema for representing quantities, as these are represented distinctly in part-whole relations between the quantities. Riley et al.'s development of these models wherein problem schemata, action schemata, and strategic knowledge determine the child's performance in word problems implies that the solution to these problems requires more than just knowing the operations and having some general skill in applying them. According to these experimenters, it is the construction of these schemata and the application of this strategic knowledge that allows a child to successfully solve word problems. They conclude that if one of these factors is missing s/he has difficulty in working out the solution between two unequal quantities.

Although this model is in many ways an impressive and imaginative one, it is partly based on an unproven assumption about the difference between Equalize and Compare problems, as already seen. The best thing to say at this stage is that the model predicts a difference between these two problems: it predicts that Compare problems will be harder than Equalize ones. It also predicts that the difference between Equalize and Compare problems would apply equally to continuous and to discontinuous quantities.

1.7.1.2. The Briars and Larkin Model

Briars and Larkin (1984) produced a structural model of a different kind: it is different in that it stresses the role of counting. They argued that that solving of mathematical word problems links the ability to execute mathematical operators and the ability to apply these operators in real-world situations. When given word problems, children are asked to use mathematical knowledge to describe situations. Briars and Larkin (1984), like Riley, Greeno, and Heller (1983), emphasized the importance of whether the problems were described as static or active situations, and of whether they were described as increasing or decreasing. This is why their hypothesis is classified as a structural hypothesis here.

Their model also differentiates itself from other models in that it assumes that when young children solve mathematical word problems by using concrete materials, they do not form a mental schema to represent the information for simple problems, as the concrete representation of the problem information constructed "on the table" serves this function. Briars and Larkin based their word problem theory on an information processing view of cognitive psychology. The model involves ideas about "counters" and "re-representation".

(1) Single-role counters

These are used in the following type of situation: A child is given a word problem beginning with a statement such as, *"Jan had twelve marbles."*. The child counts out a set of twelve marbles. When hearing, *"She lost four of them."*, the child removes four marbles from the mentioned set. When the child is asked, *"How many marbles did she have left?"*, the child identifies the remaining set as "Those she had left". Hence, this remaining set is counted and the number the child comes up with after counting this remaining set is the answer given.

Examples of these single-role counters are the following:

a. A+B=?

Joe has 5 marbles. Then Tom gave him 3 more marbles. How many marbles does Joe have now? (Riley,1979)

b. A-B=?

Fred had 11 pieces of candy. He gave 7 pieces to Linda. How many pieces of candy did Fred have left? (Carpenter and Moser,1982)

c. A-?=B

Joe had 8 marbles. Then he gave some marbles to Tom. Now Joe has 3 marbles. How many marbles did he give to Tom? (Riley,1979)

(2) Double-role counters

Double-role counters are used in the problem:

a. A+?=B

"Mary has three marbles. She takes some more from a bag. She now has eight. How many did she take from the bag?" (Steffe and Johnson,1971)

"The ones taken from the bag" is a set whose numerical size is not immediately specified so that it is impossible to immediately represent this set with a collection of objects. The problem is therefore quite difficult. Briars and Larkin argued that this set is virtually ignored initially and that the child goes on to the problem of representing the next set mentioned, which is after all the set with which Mary ends up. Mary's set is represented by increasing Mary's first set to eight. Briars and Larkin use this term double-role to apply to those chips that were in the bag initially and are also now in Mary's overall set. The child is assumed to have somehow mentally tagged all the marbles in the bag, in order to then attach a second tag to all the marbles currently in Mary's set and then count all the marbles that have both these tags.

(3) Re-representing problems

Problems begin with an unknown set, such as for example,

a. ?+A=B

"Bill had some marbles. Susan gave him three more marbles. Now he has eight marbles altogether. How many marbles did Bill have to begin with?" (Hiebert,1981)

There are two ways of solving this type of problem. One of the ways is to realize that the problem can be restated. Hence C+A=B can be restated as C=B-A. These formulae tell us that the answer is obtained by taking the number of set A from the number of set B. The second way is to realize that the problem can be converted into a double-role problem by switching the first two sets around, giving A+C=B, which is the same as the double-role problem discussed above.

Another example of the re-representation problem is the following:

b. ?-A=B

Jane has some buttons. She lost 2 buttons. Then she had 7 buttons. How many buttons did Jane have to begin with? (Tamburino,1980)

The Briars and Larkin model is largely based on Change problems, but the distinction that it makes between action and static word problems leads to a clear prediction that Equalize problems which involve action changes should be easier than Compare problems which do not. Moreover its use of the notion of "counters" leads it to a definite prediction about the use of continuous and discontinuous material. The model concerns judgements about discontinuous material only and, therefore, the difficulty that it assigns to Compare problems should apply to discontinuous material only.

1.7.1.3 The Silver and Thompson Hypothesis about the Number of Steps in a Problem

Another hypothesis involving structural aspects is that of Silver and Thompson (1984). Their hypothesis is about word problems in general and does not deal specifically with the possibility of a difference between

Compare and Equalize word problems. However, their hypothesis can be applied to this difference.

Silver and Thompson (1984) claim that data from the NAEP (National Assessment of Education Progress mathematics assessments in the United States of America), and from Carpenter et al. (1980,1984), show that children have more difficulty with problems involving multi-step, as they require the application of more than one operation, than those involving only one-step, as they can be solved by a routine application of a single arithmetic operation. Taking their hypothesis a step further, and applying it to the difference between Equalize and Compare word problems, the following deduction can be made.

An Equalize problem involves an action relationship where a transformation occurs by increasing or decreasing a quantity so as to make it equal to the other. This transformation seems to involve a one-step process to arrive at the solution. By just knowing the amounts of the two quantities involved, a child can increase or decrease one quantity to make it equal to the other, hence arriving at the solution or the difference between the two quantities.

A Compare problem, on the other hand, involves a static relationship where no transformation occurs between the quantities. This type of problem seems to require a multi-step process to arrive at the solution. The child has to know first how much there is in the start-set quantity. The child then has to know how much there is in the compare-set quantity. Then, the child has to infer how much the start-set quantity was initially (on the basis of the compare-set quantity), in order to decide a change and arrive at a solution (the result-set quantity).

In other words, for an Equalize-type problem, the child may only use the start-set and the result-set to arrive at the solution (the change-set), but for a Compare-type problem, the child has to refer back to an initial state of

the start-set in order to see what the difference between the two quantities actually is.

One may draw a parallel with recent work on causal reasoning in children by DasGupta and Bryant (1989). When young children are given a sequence

Broken Cup ? Wet Broken Cup

and asked what caused the wet broken cup, young children are unable to solve this because they do not refer to the original state of the object (i.e. dry). It is possible that in the same way, 5-to-6-year-olds are unable to refer back to an original starting point which is not actually represented. It seems that children do not have to keep in mind representations for Equalize word problems, but need to have the ability to infer representations of an initial quantity and keep them in mind for a Compare word problem.

Thus it can be seen that the Silver and Thompson analysis also leads to the prediction that Equalize problems should be much easier for children than Compare ones and that the difference between Equalize and Compare problems would apply both to continuous, as well as discontinuous quantities.

1.7.1.4 Conclusions about Structural Hypotheses

There are three hypotheses about the difficulty of Compare problems. Riley, Greeno, and Heller's hypothesis stresses action word problems and static word problems, and its answer to the two questions that we posed are:

(1) that Equalize problems should be easier than Compare problems,

(2) that the difference should be as great with continuous as with discontinuous material.

Briars and Larkin's model stresses the role of hypothetical "counters" and predicts:

(1) that Equalize problems should be easier than Compare problems,

(2) that the difference should occur with discontinuous, but not with continuous material.

Silver and Thompson stress the number of steps and predicts:

(1) that Equalize problems should be easier than Compare problems,

(2) that the difference should be as great with continuous as with discontinuous material.

Thus all three theories make the same prediction about Equalize and Compare problems. However, they disagree about the effect of changing the material to continuous material. None of the theorists has satisfactorily tested his predictions about these two questions. We do not have a satisfactory comparison between Equalize and Compare problems. Nor has there been any attempt to compare children's performance in problems about continuous material to their performance in problems about discontinuous material.

1.7.2 Linguistic Hypotheses about the Difficulty of Compare Word Problems

Structural theories predict linguistic changes should have little or no effect on the task, so long as the task structure remains the same. On the other hand, there have been arguments that linguistic factors, and particularly the use of comparative terms, like "more", are extremely important in word problem solving. The fact that children find Compare word problems difficult to solve is perhaps partly due to their linguistic factors. This hypothesis, therefore, yields two entirely different predictions from structural hypotheses. One is that linguistic changes, such as introducing Compare problems without comparative terms, should make the task much easier. The second is that

Equalize and Compare problems both employ comparative terms, and therefore, there should be no difference in difficulty between the two.

Hudson's (1983) explanation for the difficulty of Compare problems fastened on the word "more". He wanted to investigate the difficulty that children have with Compare problems. He presented 64 4-to-8-year-olds with two kinds of Compare problems. Eight pictures were presented to the child, showing for example, five birds and four worms. Two questions, separated by a short interval, were asked about the pictures. The first question was: *"How many more birds than worms are there?"*. The second question was: *"Suppose the birds all race over and each one tries to get a worm!; How many birds won't get a worm?"*.

In the "Won't get"-type questions, 100% of the children responded correctly; whereas, only 64% of the children responded correctly to the "More"-type questions. This result demonstrated that the 5-year-olds were able to use their knowledge of correspondence to determine exact numerical differences between disjoint sets. Children used a one-to-one correspondence strategy for the latter question in order to arrive at a solution, and obtained the correct answer more frequently.

Hudson concluded from his results that children's errors in solving word problems are not due to a lack of quantitative actions or procedures necessary to perform a solution, but to an occasional failure to integrate computation into the total solution of a word problem. This ability was previously cited by Ballew and Cunningham (1982) as being essential to successful problem-solving. Children's success in the "Won't get" questions, mentioned above, indicates that they are skilled at establishing correspondences and determining exact numerical differences between disjoint sets. Children's unsuccessful performance with questions presented in the "More"-type form reflects their misinterpretation or inadequate comprehension of comparative constructions of the general form, "How

many...(comparative term)...than...?". Hudson concluded that linguistic difficulties account for mistakes in Compare problems, particularly difficulties with "More"-type questions. This study will be evaluated after mention of a similar study.

Later, Hudson replicated this experiment with 30 4-to-5-year-olds. He wanted to investigate children's solution strategies in answering the "Won't get"-type questions. Children were observed to use a counting strategy rather than a perceptually guided pairing strategy, as previously demonstrated.

Hudson's study deals only with Compare-type problems. His study gives an indication that the difficulty children have with word problems of this nature, may be due to a lack of understanding of the linguistic component "more". This explanation is a weak one and not necessarily valid. After all, the same comparative word is present in both Equalize and Compare word problems.

As he presented two types of Compare problems, children should have encountered difficulty with both type problems, as Compare-type problems are generally difficult for children to solve (i.e. Riley [1981]; Riley, Greeno,and Heller [1983]). The fact that the children, in Hudson's study, did not encounter any difficulty with the "Won't get"-type question, may have been due to the type of structural operations that it triggered. The structural characteristics of the "Won't get"-type question encompassed an action relationship. The structural characteristics of a Compare-type question encompassed a static relationship.

1.7.2.1 Conclusions about Linguistic Hypotheses

The only linguistic hypothesis about Compare problems (Hudson, 1983) predicts:

(1) that there should be no difference between Equalize and Compare problems because both use the difficult term "more",

(2) that there should be no difference between problems using continuous and discontinuous material.

1.7.3 A Hypothesis which Combines Structural and Linguistic Factors

In this section, this author's hypothesis which involves both structural and linguistic factors to explain the difficulty of Compare problems will be presented. This hypothesis also predicts that Equalize problems should be easier than Compare problems. The hypothesis rests on a distinction originally made by Moore and Frye (1986) between two senses of the word "more". One sense is "sequential", and the other "simultaneous". They claim that the "simultaneous" sense is the harder of the two for children. This is relevant to Compare problems because Compare problems use the word "more" in the simultaneous sense. Equalize problems, on the other hand, use it in the "sequential" sense, which according to Moore and Frye, is the easier sense. Thus, their analysis predicts that Equalize problems should be easier than Compare problems.

Moore and Frye (1986) did some work on children's performance on quantity tasks relating to children's understanding of the concept, "more". They noted that a "sequential" as well as a "simultaneous" meaning is derived from this particular word. The sequential meaning refers to an addition over and above the amount already present. This sequential meaning can be attained as early as the age of two when the child begins to understand terms such as "more milk". In terms of word problems, "sequential" is semantically synonymous with "Equalize", where a quantity is increased or decreased. The simultaneous meaning refers to "more" as being the greater of two or more different quantities. In terms of word problems, "simultaneous" is

semantically synonymous with "Compare", where two distinct quantities are compared.

Moore and Frye postulated developmental changes of the concept "more". They hypothesized that children undergo a process whereby the sequential meaning of "more" is the one which makes sense to them, followed by a period where both the sequential and the simultaneous meaning are integrated jointly, and finally by the process where the simultaneous meaning is understood. Children under 5-years-of-age are not able to understand the meaning of "more" in the simultaneous sense, but are able to understand the meaning of "more" in a sequential sense. The use of "more" in its sequential form is often said to be acquired within the first two years, as in "more milk" (Bloom,1970). Hudson et al (1982) also found that when water was poured from a larger amount to a lesser amount, still leaving the former greater, children tended to choose the larger amount as having more. Although the children would presumably understand "more" to mean an addition, very few applied this meaning to the context of this task. Hence, there is evidence (Bloom,1970) that the sequential meaning of "more" is developmentally prior to the simultaneous meaning, but that the latter comes to dominate in quantification tasks (Hudson et al., 1982).

Moore and Frye tested one hundred and forty-four children of ages 4-, 6-, and 8-years-old, in a task where the term "more" could be given different meanings. They were presented with four types of problem which involved quantity comparisons.

(1) A standard conservation of continuous quantity, in liquid form, where one of two same-sized jars contained more water than the other. The water from the jar with less was then poured into a taller, thinner jar so that the level rose above that in the comparison jar.

(2) An addition task where one of two same-sized jars contained more than the other. A small fixed amount was poured from the jar with more into

the jar with less. This transformation left the jar that had previously contained less with a still noticeable deficit.

(3) A trick task where the container utilized, a jar, had a false bottom, with the water level higher than in the normal jar, but where, in fact, the former jar contained less than the latter. The transformation consisted in pouring the water from the trick jar into a new normal jar at the same time as some additional water from a fourth jar. This transformation resulted in the original normal jar having a higher level than the new jar.

(4) A fixed task with no transformation, where the two same-sized jars resembled the end display of the addition task. Children were asked, *"Which has more?"*.

It was hypothesized that,

(1) The 4-year-olds would mostly interpret the meaning of "more" in a sequential sense, as they would be less likely to judge by appearance, especially on tasks involving an addition (i.e. they would judge according to addition).

(2) The 6-year-olds would interpret the meaning of "more" in both the sequential and simultaneous sense, as they are non-conservers and would tend to judge by appearance in all tasks, just as they interpret "more" simultaneously in tasks in the standard context.

(3) The 8-year-olds would apply the simultaneous meaning most frequently, as they are within the conservation age range, and would be expected to succeed on the shape change task, as well as show a greater flexibility in their response strategies.

The following results were observed. In the addition task, the 4-year-olds were less likely to base their judgement on appearance and interpreted "more" in the sequential sense. In the fixed task, however, this younger group showed that they could interpret "more" in the simultaneous sense. The 6-year-olds judged according to appearance on all tasks, and used

height as the basis of their judgement. Some of the 8-year-olds, who were conservers, interpreted "more" in the simultaneous sense and thus judged according to appearance. In the addition task, they chose the highest quantity rather than the quantity that had been added to. On the other hand, it was observed that 8-year-olds are capable of applying the sequential meaning, as they performed correctly on the trick task by choosing the jar with the lower level as "more". This indicated, they believed, that if there had been an addition to an amount that was previously greater than another amount, then the former must still be greater. Appearance was not relevant here. The results support the hypothesis that in quantity tasks the child's understanding of the word "more" changes developmentally.

Moore and Frye's findings have an important bearing on Compare and Equalize problems. They indicate that the outcome of these stories which present a mathematical problem might depend on the understanding of the word "more". This understanding, according to Moore and Frye's findings, changes developmentally. Thus, as a child increases in age so does his/her capacity to compare unequal quantities and solve word problems.

Moore and Frye claim that children's difficulty in solving problems is due to a linguistic aspect. However, children's difficulty in their task may have been due to a more basic problem, rather than to a linguistic one, as they suggest. Children, in their task, may have been misguided by these perceptual transformations and hence, performed as they did. The fact that 6-year-olds judged according to height only proves this point. Had these perceptual transformations not been presented, children may have performed differently. It is unclear whether the problem in Moore and Frye's task is actually more fundamental than linguistic.

1.7.3.1 Conclusions about the Hypothesis which Combines Structural and Linguistic Factors

This hypothesis predicts:

(1) that Equalize problems should be easier than Compare problems,

(2) that the difference should be as great with continuous as with discontinuous material.

1.8 General Conclusions

The structural aspects of the word "more" have to be taken into account. Hence, the Equalize/Compare difference can be explained by an amalgamation or combination of the structural properties and the linguistic properties of word problems.

Combination of Structural and Linguistic Factors for Equalize Word Problems	**=**	**Action x Sequential**	**=**	**"Within"-Type Relation**
Combination of Structural and Linguistic Factors for Compare Word Problems	**=**	**Static x Simultaneous**	**=**	**"Between"-Type Relation**

Figure 6. Visual representation of this author's hypothesis as an explanation to the Equalize/Compare difference.

The above graphical figure of the hypothesis combining both structural and linguistic factors shows that in an Equalize-type problem, "sequential" is

semantically synonymous with "equalize", in terms of word problems, where a quantity is increased or decreased. A child comprehending "more" in this particular linguistic framework, understands, structurally, that a transformation of action relationship is taking place, as a comparison is made of the same quantity in two different states. This structural and linguistic combination for an Equalize-type question is classified in a "within"-type relation.

On the other hand, the above graphical figure of the hypothesis combining both structural and linguistic factors shows that in an Compare-type problem, "simultaneous" is semantically synonymous with "compare", in terms of word problems, where two distinct quantities are compared. A child comprehending "more" in this particular linguistic framework, understands, structurally, that a static relationship or no transformation is taking place. This structural and linguistic combination for a Compare-type question is classified in a "between"-type relation.

One way to investigate this model is to test its prediction that Compare problems should be more difficult than Equalize ones. The other main way is to attempt to answer a basic question which should have been asked from the beginning. To what extent is children's performance in word problem solving specific to number (or discontinuous quantity)? Would the same factors apply in problems which involve continuous material and non-identifiable numbers at all?

Aims of the thesis:

The experiments in this thesis will have 5 main aims.

1. To compare children's performance in Equalize and Compare tasks;
2. To look at the effect on performance in these two kinds of problem of continuous versus discontinuous material;
3. To look at children's counting strategies in problems involving discontinuous material;
4. To look at children's strategies in problems involving continuous material;
5. To look at the effect of varying the comparative terms used.

CHAPTER TWO

CHILDREN'S UNDERSTANDING OF MATHEMATICAL WORD PROBLEMS

2 Experiment 1 An Investigation of Responses made in Different Types of Question using Different Types of Material Depicting Quantity

2.1 Introduction

This first experiment investigates children's understanding of Equalize and Compare word problems and their ability to make use of graphical information in such problems. Several studies have been carried out concerning children's difficulty with Compare questions. However, Equalize questions have been relatively neglected. These questions involve the same mathematical calculations presented in different ways. An Equalize-type question is of the form, *"How much more does B need to have, to have the same as A?"*, while a Compare-type question is of the form, *"How much more does A have than B?"*.

These earlier findings raise questions as to the nature and causes of the difficulty that children experience with Compare-type questions and in particular, whether there is a difference between Equalize and Compare questions. If there is a difference, one also needs to know whether it is as great with continuous material, as with discontinuous material (see Chapter One). All the previous experiments dealt only with discontinuous material. Hence, discontinuous and continuous material were presented to the children, as well as verbal discontinuous questions.

2.2 Method

(a) Subjects

Two groups of 18 children served as subjects. One ranged in age from 5;4 to 6;8 (mean age of 5;9), the other from 6;8 to 8;2 (mean age of 7;3). An additional group of 14 children (age-range 5;3 to 7;11; mean age of 6;9) served as pilot subjects. Children serving as subjects were chosen according to their performance on two standardized tests: the short form of the British Picture Vocabulary Scale (BPVS) and the arithmetic section of the Weschler Intelligence Scale for Children Revised (WISC-R), as well as a test devised for this experiment, including four addition problems and four subtraction problems (see Appendix 3). Those children who obtained the highest scores were chosen as subjects. Those who obtained lower scores served as pilot subjects. This was done so as to avoid the risk of low-scoring children reaching a floor effect in the task and not being able to do the experiment, which might have confounded results relating to differences between conditions. The children were all attending a First School in Oxford, England.

(b) Design

A mixed design with repeated measures on Conditions (2: Question Type [Q], Material [M]) was used. Age (2: 5, 7) was the Between factor.

Three sets of sequences (I:A-B-C, II:C-A-B, III:B-C-A) were created for material display. Condition A was the continuous condition, Condition B was the verbal discontinuous condition, where no pictorial material was presented to the child), and Condition C was the discontinuous condition. Three sub-groups of six children in each group experienced a sequence-set (I,II,or III). This was done so as to control for order effects.

Again, to control for order effects, subgroups of three children were formed from each group of six children in each sequence-set. A sub-group of three children was presented with the Equalize question first and the Compare question second (sequence set X); the other group of three children was presented with the Compare question first and the Equalize question second (sequence set Y). Hence, for each condition (A,B,or C) half the children received the Equalize question first and the Compare second; the other half received the Compare question first and the Equalize second. Thus, all the children performed under both conditions and experienced both sets of sequences, but each group of three children experienced a different order of presentation for each condition, as illustrated in Table 2.2.1.

(c) Materials

The stimuli consisted of illustrations representing quantities in either continuous or discontinuous fashion (see Figures 2.2.1 and 2.2.2). The continuous material consisted of a series of four 9 x 13 inches illustrations. Each illustration showed two sets of items. The first set of items consisted of two block rectangles pasted on the left-hand side of the illustration. The left-hand rectangle was always 7 inches high and had a girl's face above it. The rectangle next to it was of different heights in different illustrations: either 6,5,4,or 3 inches. A boy's face was above it.

The second set of items consisted of five block rectangles, pasted on the right-hand side of the illustration. The first was always 7 inches in height. The other four rectangles were taken randomly from a set of 6 rectangles that were 6,5,4,3,2, and 1 inch in height. The set of items on the left-hand side were always a subset of those on the right-hand side. The items in each of the two sets were always positioned in descending order, from tallest to shortest, with the tallest situated on the left (Hudson,1983). Each of the 4

illustrations was presented in a different colour: black, orange, red, or yellow. All rectangles in each set were positioned 1 inch apart from each other and there was a gap of 2 inches between the two sets (see Figure 2.2.1).

The discontinuous material was similar to the continuous, except that the numerical quantities were here represented by small circles. The same design was used as for the continuous material, except that, for example, a 7 inch rectangle was replaced by seven small circles. The height in inches of the rectangle in the continuous condition always corresponded to the number of circles in the discontinuous condition (see Figure 2.2.2).

(d) Procedure

Each child was tested individually under three conditions: Condition A (the continuous condition), Condition B (the verbal discontinuous condition, where no pictorial material was presented to the child), and Condition C (the discontinuous condition). The order of presentation of the conditions was counterbalanced between children in a 3x3 Latin Square design.

In each condition, each child was asked eight questions. There were two types of question: Compare and Equalize. Four Compare questions and four Equalize questions were assigned to each condition. Equalize and Compare questions were typically blocked across conditions, with order of blocks counterbalanced across subjects.[2] The children were grouped in each set of conditions according to their performance on the previously mentioned tests. This was done so as to avoid the risk of the best-scoring children all being given the same condition, which might have confounded results relating to differences between conditions.

So, the two orders of presentation were (1) Equalize question on first trial and Compare question on second; (2) Compare question on first trial and Equalize question on second. The material groups were (1) set A-B-C for

[2] This aspect of procedure is repeated in the following Experiments.

one group of 6 children; (2) set C-A-B for another group of 6 children and; (3) set B-C-A for a final group of 6 children.

The phrasing of a Compare question in the continuous task was as follows, *"This girl has this much chocolate; this boy has this much chocolate. How much more chocolate does the girl have than the boy?"*. Equalize questions were phrased differently. The experimenter said, *"How much more chocolate does the boy need to have, to have the same as the girl?"*. For the discontinuous material, the Compare question was, *"This girl has so many sweets; this boy has so many sweets. How many more sweets does the girl have than the boy?"*. Equalize questions in this condition were phrased as follows, *"How many more sweets does the boy need to have, to have the same as the girl?"*. In the verbal discontinuous condition, no pictorial material was included to help the child. The questions asked were identical to those used in the discontinuous task, except that the preliminary introduction included real numbers. An example of a Compare question in this condition was: *"Jane has seven sweets and Bob has six. How many more sweets does Jane have than Bob?"*.

To summarize, each illustration presented consisted of a pictorial version of the question asked (set 1) and a set of five possible responses (set 2). In all tasks, if the child responded verbally to the question, s/he was encouraged to point to the correct answer on the right-hand side of the illustration.

2.3 Results

The principal data were the number of Equalize and Compare questions answered correctly, in each of the three types of material. The mean number of correct responses for each type of question under each material type is shown in Table 2.3.1. These means were obtained from

scores on the four questions on each question type under each type of material and give an overall picture of children's performance. In general more correct responses were obtained from the Equalize questions than from the Compare questions. However, this Equalize and Compare difference was not evident in the older age group. In both questions, 7-year-olds made more correct responses than 5-year-olds. This Equalize and Compare difference looks more striking in the younger group. In general, the difference between Equalize and Compare questions seems to be the same for all three types of material.

Table 2.3.1 **Means and Standard Deviations (out of 4) for the Number of Correct Responses**

2.3.1 Comparing Correct Scores across the Two Ages

The means and standard deviations for the total number of correct responses made in each condition by the two age groups were given in Table 2.3.1. This table shows that, first, there was a difference between Equalize and Compare questions, where Equalize questions were easier than Compare questions for the younger children. There was little or no difference for the older children because of a ceiling effect. Second, this difference between Equalize and Compare questions seems to be as great with continuous as with discontinuous material.

2.3.2 Analysis of Variance

These trends were analyzed by subjecting the raw scores for the total number of correct responses to a 2x3x2 Analysis of Variance (henceforth referred to as ANOVA) with repeated measures. There was one between-subject variable: Age-group (younger or older). There were two within-subject variable: Question (Equalize or Compare) and Material

(discontinuous, continuous, or verbal discontinuous). The results of this analysis are presented in Table 2.3.2.

The Age term was statistically significant ($F(1,34)=38.13$, $p<0.001$). This shows that there was a tendency for older children to make more correct responses than the younger ones.

The significant main effect of Question ($F(1,34)=13.13$, $p=0.001$) indicated that children made significantly more correct responses to Equalize questions than to Compare questions.

The significant main effect for Material ($F(2,68)=4.17$, $p<0.05$) shows that children made significantly more correct responses when presented with discontinuous material, than when presented with continuous or verbal discontinuous material.

Graphical representations of these main effects are presented in Figure 2.3.1.

The significant interaction between Question Type and Age ($F(1,34)=8.60$, $p=0.01$) showed that the effect of Question Type varied between the two age groups. This interaction was explored using the Newman-Keuls Multiple Range Test (Bruning and Kintz, 1977:119-122). This post-hoc test showed that for the older group there was no difference between type of question whereas for the younger group, Equalize questions were significantly easier than Compare questions. Hence, there was not an Equalize and Compare difference for the older group, but there was for the younger group. A graphical representation of this comparison is presented in Figure 2.3.2.

The significant interaction between Material and Age ($F(2,68)=4.17$, $p<0.05$) showed that the effect of Material Type varied between the two age groups. This interaction was explored using the Newman-Keuls Multiple Range Test (Bruning and Kintz,1977:119-122). This post-hoc test showed that, with respect to material, the 7-year-olds were not affected by the type of

material presented, whereas with the younger group, discontinuous material was a lot easier than continuous material, and these were much easier than the verbal discontinuous material. A graphical representation of this comparison is presented in Figure 2.3.3.

There was no significant interaction between Question Type and Material ($F(2,68)=2.83$, n.s.) which indicated that the difference between Equalize and Compare problems did not vary across the three types of material. The difference between the two question types was not significantly greater with one type of material than with the other.

The cell mean score and standard deviation for the older age group in the Equalize question when presented with discontinuous material, as presented in Table 2.3.1, showed no variance. Therefore, a separate ANOVA was performed of the younger group's raw scores for the total number of correct responses. The results of this analysis are presented in Table 2.3.3. Previous effects are confirmed.

In summary, Equalize and Compare questions are different from each other. Equalize questions prove to be easier than Compare questions, but this applies only to the younger children. Also, the Equalize and Compare difference appears to be as great with continuous as with discontinuous material.

2.4 Discussion

The ANOVA has produced an interesting set of results. First, it has answered the question as to whether there is a difference between Equalize and Compare word problems. The finding that children respond better to Equalize questions than to Compare questions supports the hypothesis that children have particular difficulty with questions that ask them to make comparisons between quantities versus those that ask them to make

comparisons within a quantity. They perform better with questions which ask them to equalize amounts, rather than those which ask them to make comparisons between quantities. However, this was only true for the younger age group. There was a ceiling effect in the older group, which meant that they made virtually no mistakes in any of the problems.

Secondly, it has answered the question as to whether there is a difference between material type. The results of Experiment 1 indicate that 5-year-old children find discontinuous material easier than continuous material. The 7-year-olds, however, reached a ceiling effect in performance, indicating that by this age level they can already deal with different types of graphical information. However, the difference between Equalize and Compare questions applies equally to both these kinds of material. This finding supports the hypothesis that the Equalize and Compare difference is not specific to number and to counting, but more fundamental, to quantity in general.

The Equalize and Compare difference had not been investigated before. This difference is mathematically a very interesting result as both Equalize and Compare word problems are basically the same mathematical problem dealing directly with how children compare two unequal quantities, and which require the same computation. Yet, children seem to have difficulty with one and not with the other. This finding indicates that although these two types of word problems may be the same mathematically, they are not the same psychologically. Children seem to approach each problem differently. It is important to find out what is the psychological basis of this difference, as when it occurs, it is not confined to discontinuous material. Therefore, the Equalize and Compare difference is due to something very general about quantity.

In the next experiment, a more systematic and controlled display was used by repeating this experiment on a BBC microcomputer. As it was the

younger children, in Experiment 1, who were significantly affected by Question Type, Experiment 2, dealt only with this 5-to-6-year-old group.

FIGURES AND TABLES FOR EXPERIMENT 1

Table 2.2.1 The Design of Experiment 1

No. of children 5-yr	No. of children 7-yr	Material Type Cond.	SS	No. of children	Question Type Cond.	SS
6	6	I	A	3	A:	x
					B:	y
			B		C:	x
			C	3	A:	y
					B:	x
					C:	y
6	6	II	C	3	C:	x
					A:	y
			A		B:	x
			B	3	C:	y
					A:	x
					B:	y
6	6	III	B	3	B:	x
					C:	y
			C		A:	x
			A	3	B:	y
					C:	x
					A:	y
N= 18	N= 18					

I = Sequence Set [SS] A-B-C
II = Sequence Set [SS] C-A-B
III = Sequence Set [SS] B-C-A

x = Sequence Set Equalize first, Compare second
y = Sequence Set Compare first, Equalize second

Figure 2.2.1 Example of Continuous Displays.

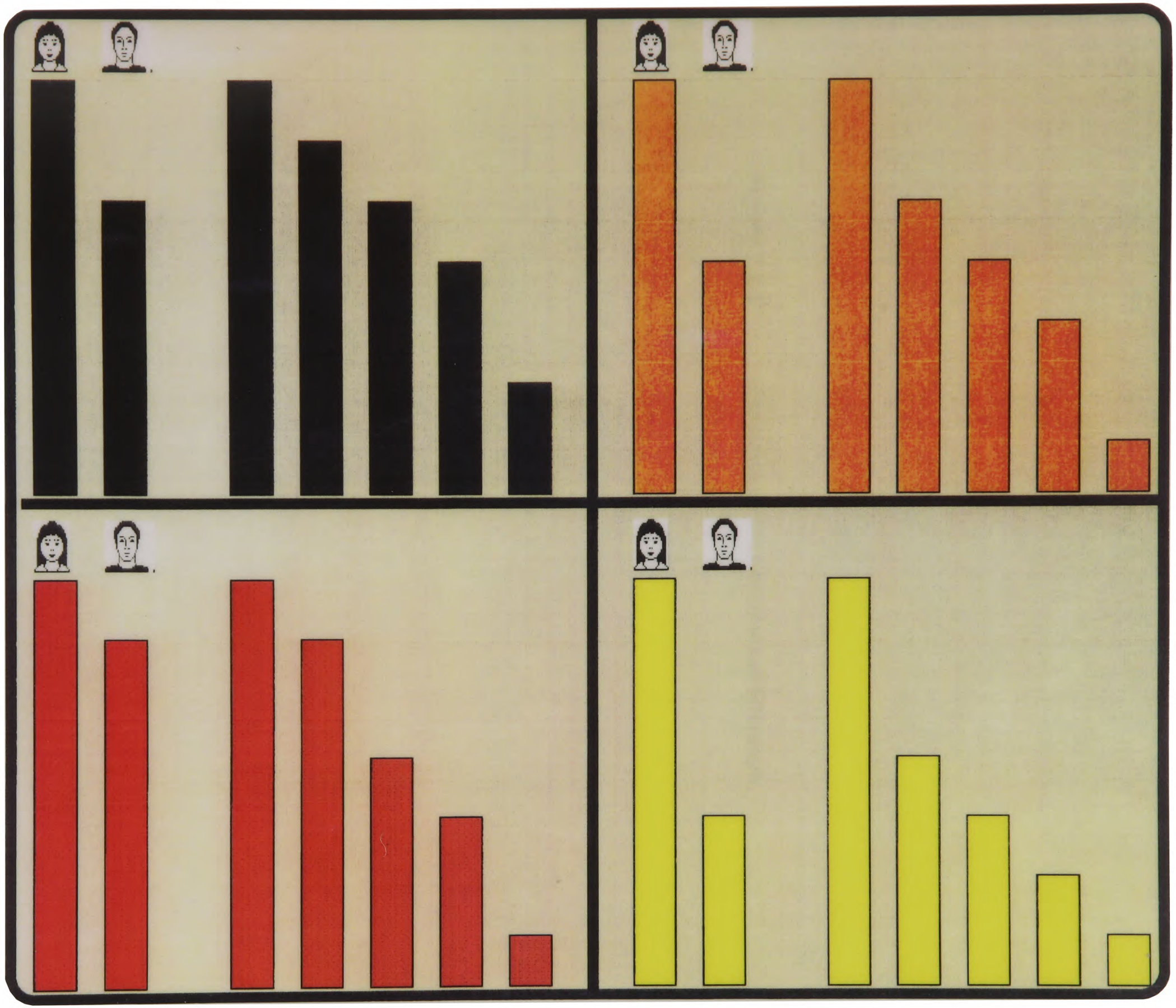


Figure 2.2.2 Example of Discontinuous Displays.

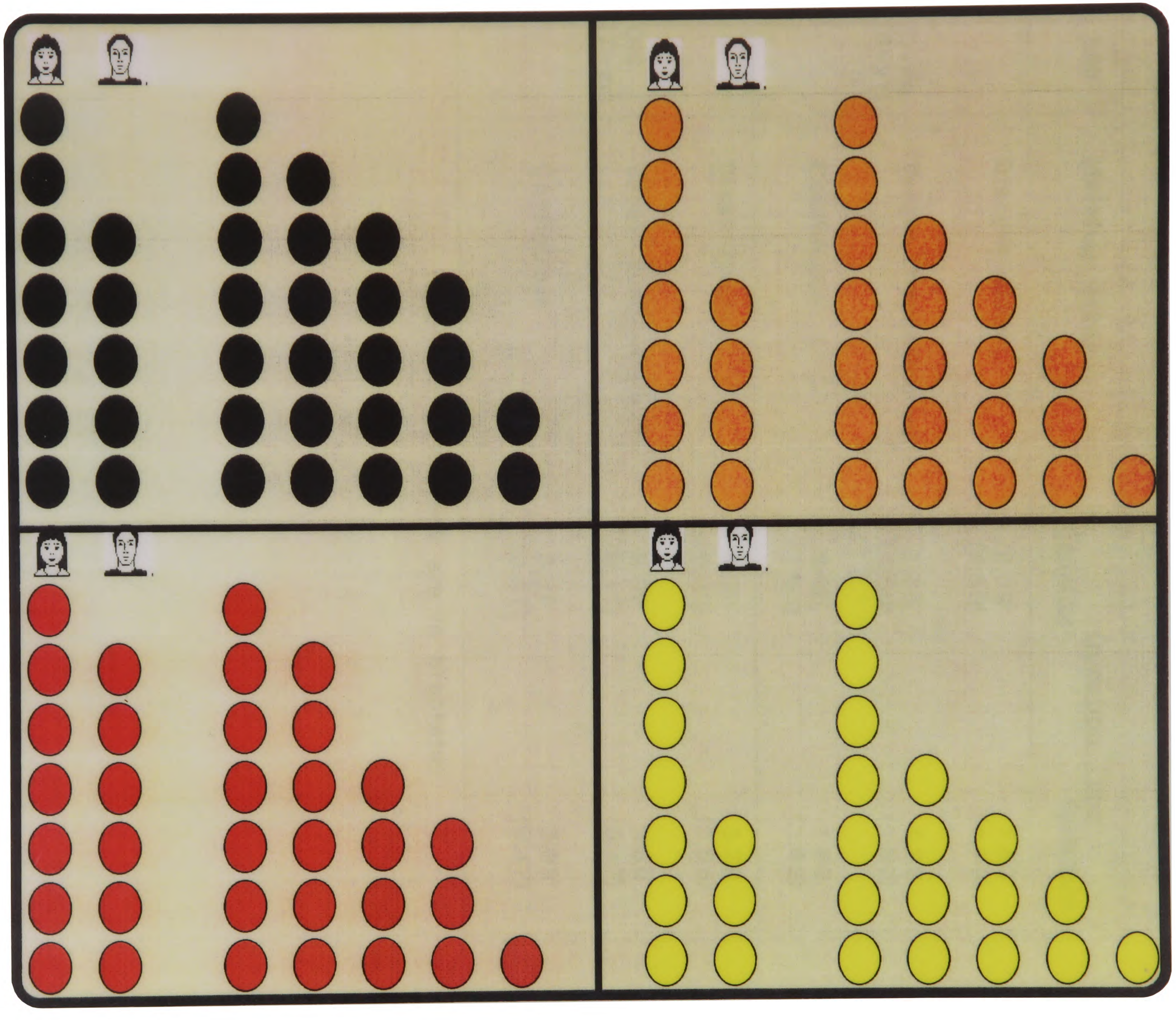


Table 2.3.1 Means and Standard Deviations (out of 4) for the Number of Correct Responses

Age	Material Group	Question Type: Equalize	Question Type: Compare
Young (5-yr)	Discontinuous	2.89 (1.53)	1.83 (1.79)
	Verbal Discontinuous	1.78 (1.59)	1.22 (1.67)
	Continuous	2.00 (1.61)	1.50 (1.62)
Old (7-yr)	Discontinuous	4.00 (0.00)	3.78 (0.65)
	Verbal Discontinuous	3.89 (0.47)	3.89 (0.47)
	Continuous	3.89 (0.47)	3.89 (0.47)

(Standard deviations are in brackets.)

Table 2.3.2 Summary Table (ANOVA) for Total Correct Scores

Source	SS	DF	MS	F	PROB.
Mean	1791.130	1	1791.130	310.41	0.0000
Age[A]	220.019	1	220.019	38.13	0.0000
error	196.185	34	5.770		
Material[M]	7.065	2	3.532	4.17	0.0195
MA	7.065	2	3.532	4.17	0.0195
error	57.537	68	0.846		
Question[Q]	8.167	1	8.167	13.13	0.0009
QA	5.352	1	5.352	8.60	0.0060
error	21.148	34	0.622		
MQ	1.694	2	0.847	2.83	0.0659
MQA	0.287	2	0.144	0.48	0.6212
error	20.352	68	0.299		

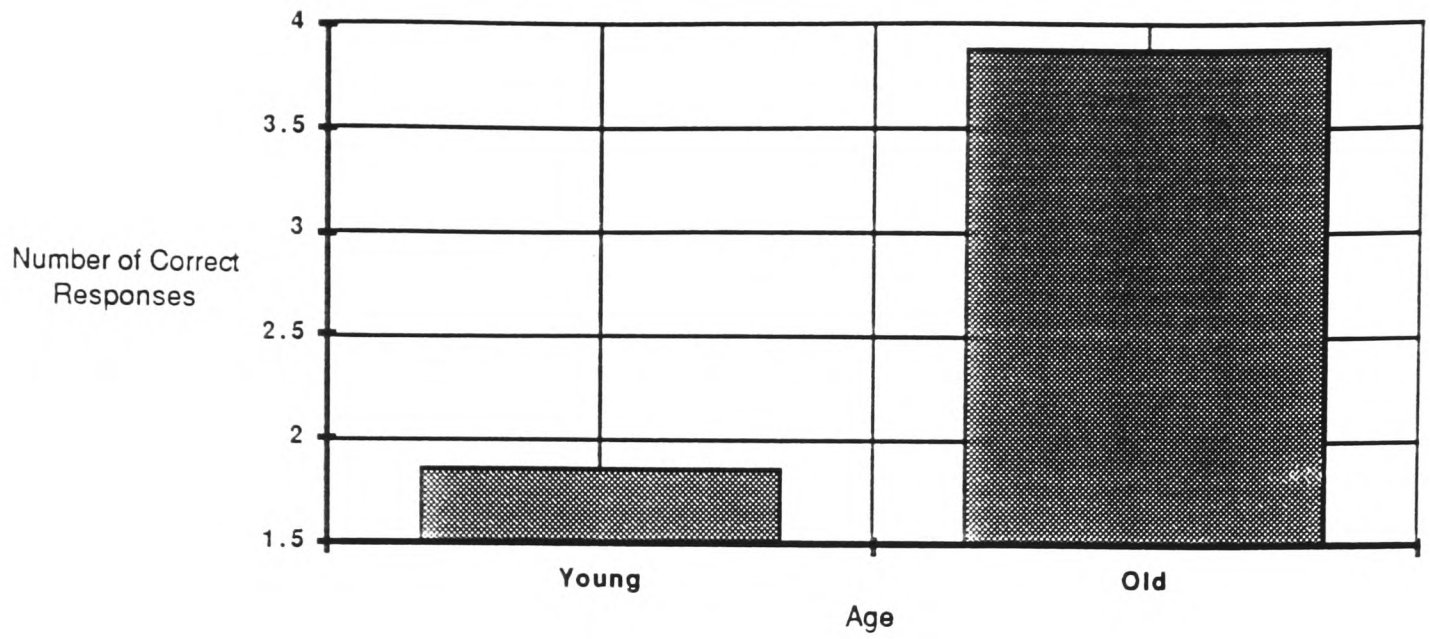
SS = Sum of Squared Deviations
DF = Degrees of Freedom
MS = Mean of Squared Deviations
F = Ratio of Variances
Prob. = Level of Significance

Table 2.3.3 Summary Table (ANOVA) for Total Correct Scores

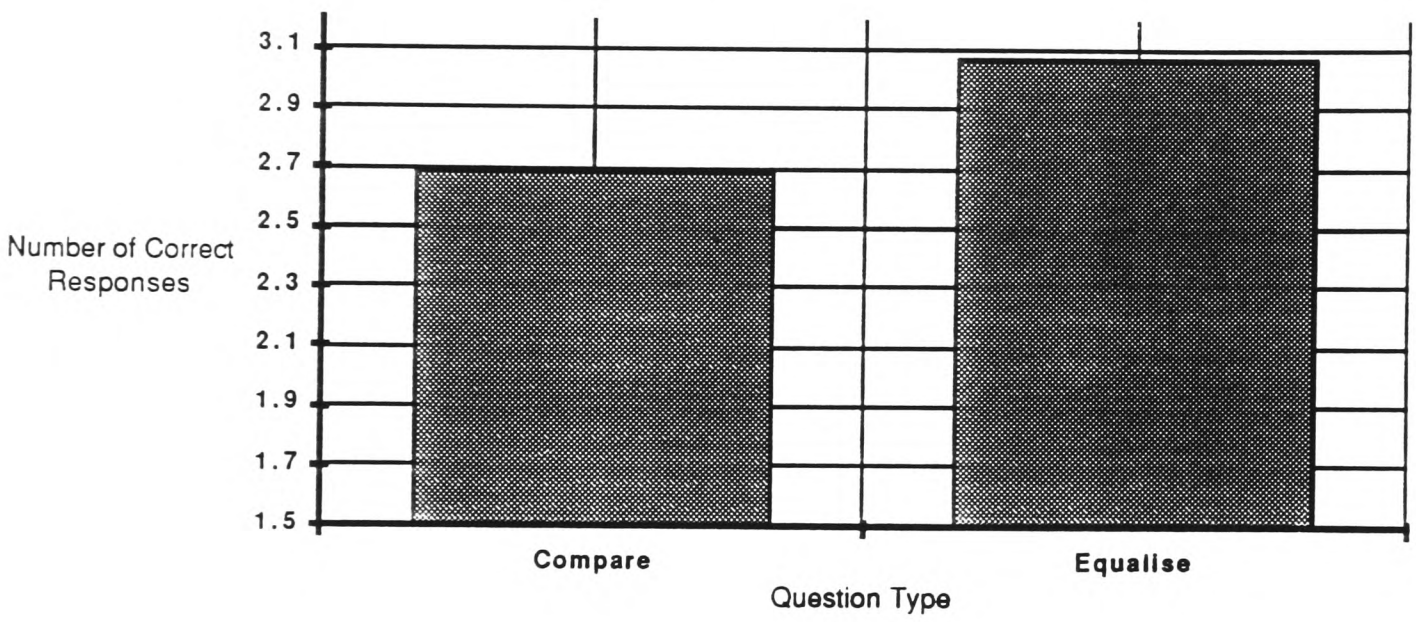
Source	SS	DF	MS	F	PROB.
Mean	377.815	1	377.815	35.65	0.0000
error	180.185	17	10.599		
Material[M]	14.130	2	7.065	4.38	0.0203
error	54.870	34	1.614		
Question[Q]	13.370	1	13.370	11.39	0.0036
error	19.963	17	1.174		
MQ	1.685	2	0.843	1.59	0.2181
error	17.982	34	0.529		

SS = Sum of Squared Deviations
DF = Degrees of Freedom
MS = Mean of Squared Deviations
F = Ratio of Variances
Prob. = Level of Significance

Comparative Performance Of Age Variable Mean Scores Based On Total Number Correct In All Tasks



Comparative Performance Of Question Variable Mean Scores Based On Total Number Correct



Comparative Performance Of Material Variable Mean Scores Based On Total Number Correct

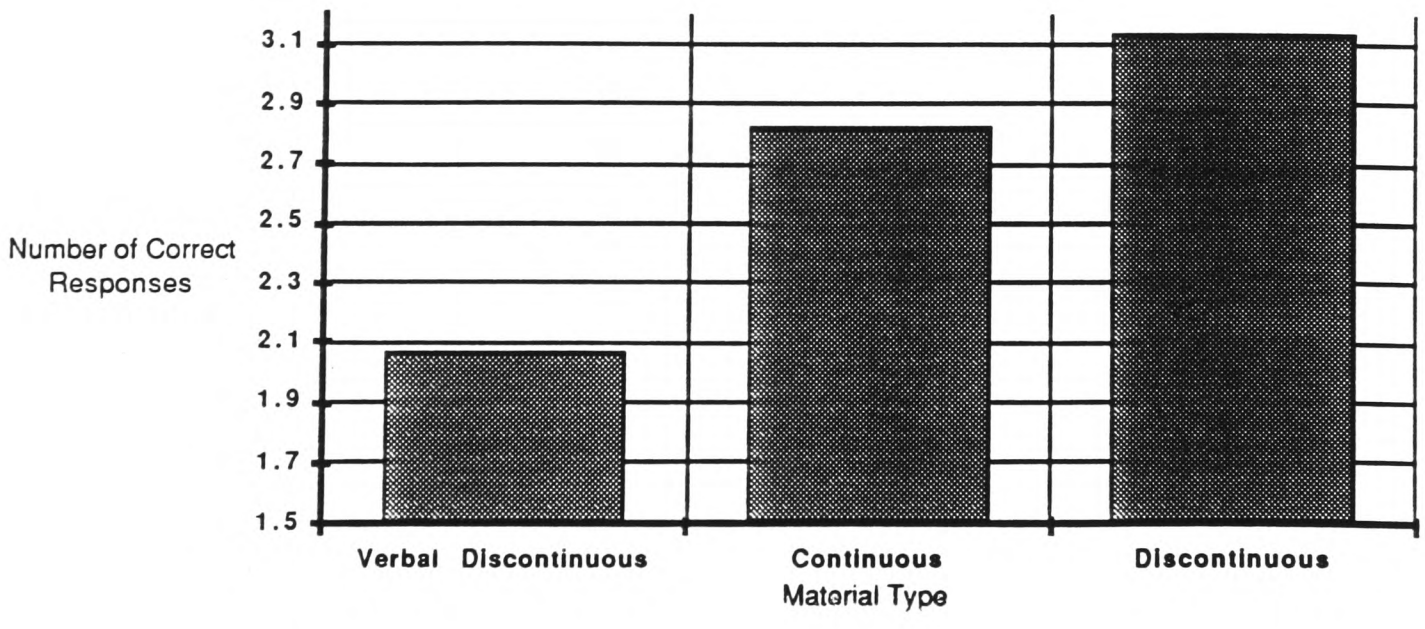


Figure 2.3.1 Graphical representations of the main effects.

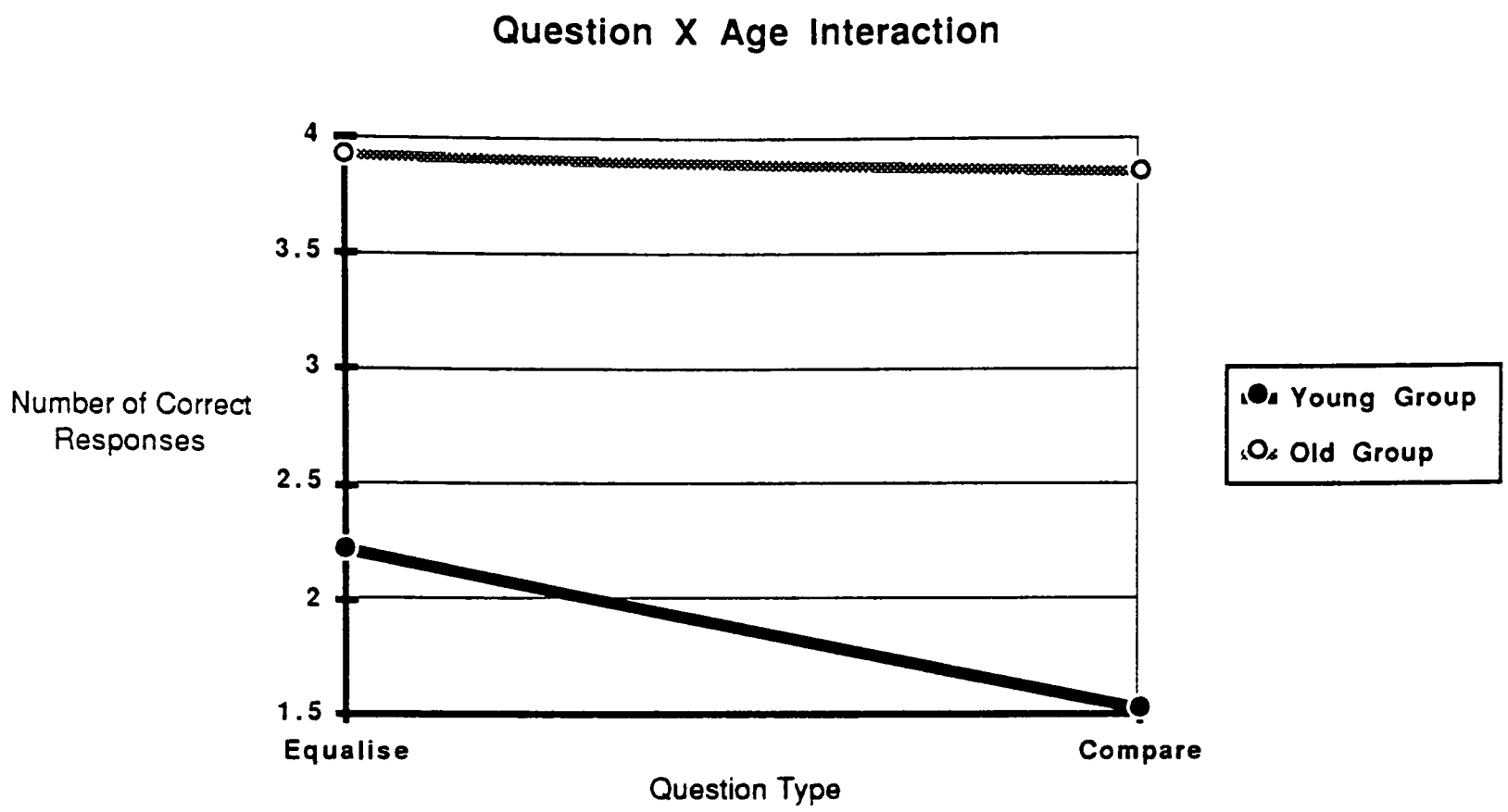


Figure 2.3.2 Graphical representation of the interaction.

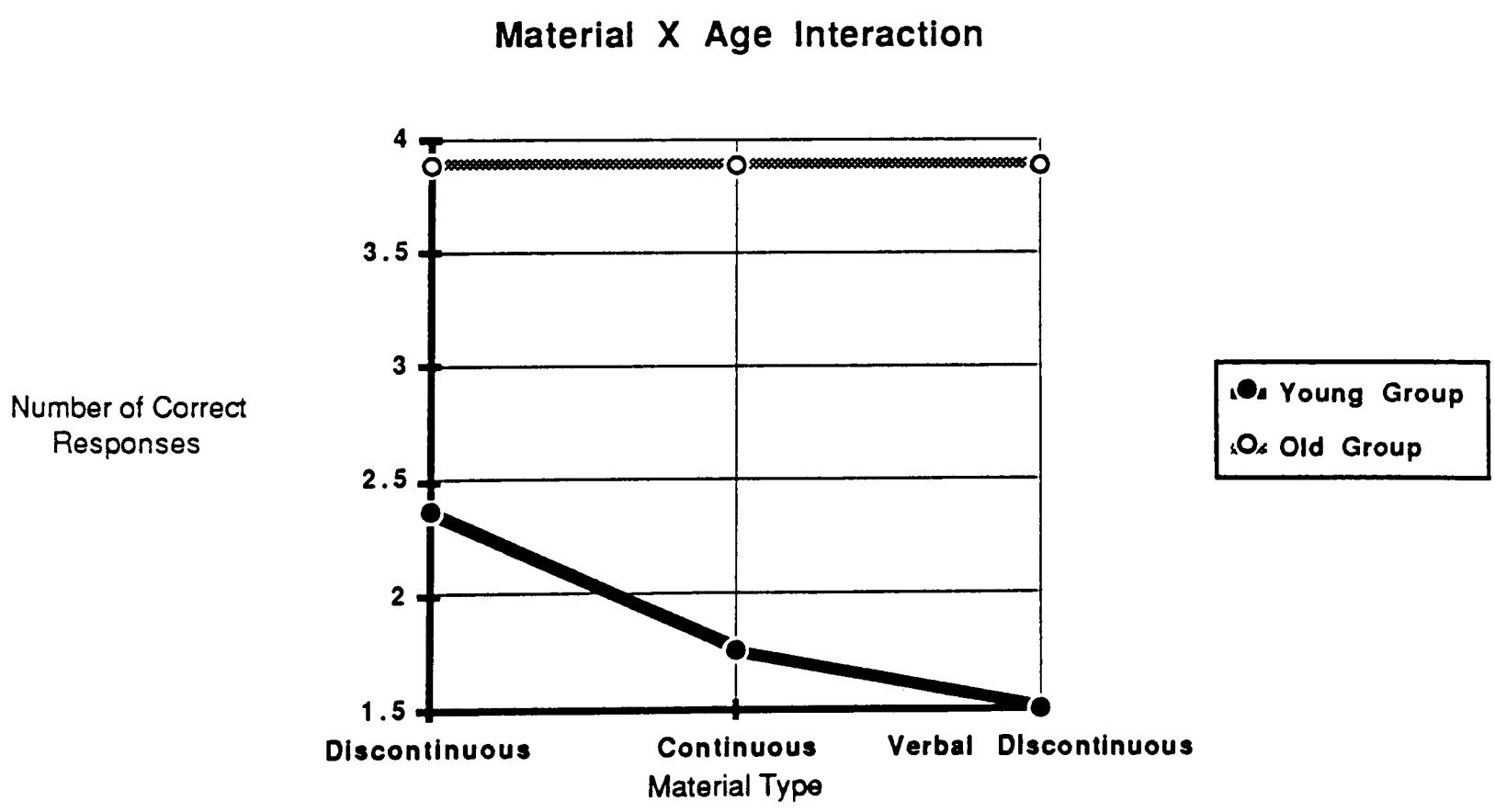


Figure 2.3.3 Graphical representation of the interaction.

CHAPTER THREE

CHILDREN'S UNDERSTANDING OF MATHEMATICAL WORD PROBLEMS ON A COMPUTER

3 Experiment 2 An Investigation of Responses made in Different Types of Question using Different Types of Material Depicting Quantity on a Computer

3.1 Introduction

Experiment 1 provided evidence, not previously reported, that young children (5-to-6-years old) still find difficulty with Compare word problems even when they are presented with continuous graphical information. In order to answer the question adequately as to whether quantitative comparisons are number-specific relating to discontinuous material or non-number-specific relating to continuous material, a more systematic and controlled method of presentation was used in Experiment 2, by repeating Experiment 1 on a BBC microcomputer. The aims of this experiment were to repeat Experiment 1, with some variations. These variations included presenting the material, as previously mentioned, using a more controlled method of presentation and introducing both comparative terms: "more" and "less".

This thesis has reviewed literature on the linguistic comparative term, "more". (See Sections 1.7.2 and 1.7.3.) There also is a vast amount of research on the comparative term, "less".

Donaldson and Balfour (1968), for example, devised a study comprehension of the words "more" and "less". They gave a group of 15

children, ages between 3;5 and 4;1, two tasks. The second task was given six months later.

The first study involved questions using "more" and "less" in different stimulus situations. 1) One situation involved making a comparative judgement with respect to a situation of static inequality. In one condition, the child was shown two blank trees and asked to put more apples on one tree than on the other; in the other condition, the child was asked, *"Does one tree have more (less) apples on it than the other?"*. If the child replied 'yes', s/he was then asked, *"Which tree has more (less) apples?"*. 2) A second situation involved making a comparative judgement with respect to a situation of static equality. The wording of the questions was the same as above, but if the subject denied that one tree had more apples than the other, s/he was further asked, *"Is there the same number of apples on each tree?"*. 3) The third situation involved a situation of initial equality which was then changed by addition or subtraction. In one condition, the child was shown the two trees with equal sets of apples, then one (or two) additional apples were held up, and the child was asked, *"If I put this apple on this tree, will one tree have more apples than the other tree?"*. If the child answered 'yes', s/he was then asked, *"Which tree will have more apples?"*. The addition was made and the child asked, *"Does one tree have more apples on it?"*. If the child said 'yes' again, s/he was asked which one. The procedure was the same for the comparative term "less", but instead, the experimenter removed one or more apples from one of the trees. In the second condition, the child was asked, *"Make it so that there are more (less) apples on this tree than on this one."*. 4) The fourth situation involved a situation of initial inequality where one or more apples were added to the tree which initially had less (or removed from the tree which initially had more). The change brought about was always too small to reverse the direction of the initial difference. Thus, the final state did not accord with the direction of change. 5) The final situation involved a

situation of initial inequality where the child was asked to reverse its direction. The instruction was worded: *"Now make it so that there are more apples on this tree* (lesser number) *than on this tree"*, or *"...so that there are less apples on this tree* (greater number) *than..."*.

For all questions, the numbers used ranged from one to five and the differences between sets ranged from one to four.

Results, for this first study, demonstrated that in the great majority of the cases, children did not differentiate "less" from "more". Questions which contained the word "less" were answered as though they were questions which contained the word "more".

In the first situation, where the children were given two blank trees and told to put more (or less) apples on one than on the other, most of the children, with the exception of three, succeeded in the task with "more". However, only two children succeeded in the task with "less"; the other children chose the tree with the greater number, as they had done previously in the question with "more". On 40 trials, the child having judged that one set contained less apples than the other, proceeded to choose, as the one that had less, the one that in fact contained more. Hence, the general finding is one where the children showed a marked tendency to interpret "less" as synonymous with "more".

In the second situation, where the children were presented with situations of initial inequality and asked whether one tree had more (or less) apples on it than the other, most children gave evidence of denial of inequality by denying that one tree had more (less) than the other. Again, some children responded to "less" in the same way as they did to "more" by using expressions like, *"On each tree"* or *"Both the trees"*.

In the third situation, where the children were required to predict which tree would have more based on addition or subtraction, by far the most common response, whether for "less" or for "more", was to add to the

indicated tree. As for previous questions, children could for the most part correctly indicate which tree had more when there was inequality, but when asked which tree had less, they again showed a marked tendency to choose the tree which had more.

In the fourth situation, where there was conflict between the act of adding (or removing) apples and the final static comparison of the two sets (i.e. when the tree to which more had been added was nevertheless the tree which had less at the end), only six children could successfully handle the task with "more". In the task with "less", seven children answered the question by choosing the tree which had more.

In the fifth and final situation, which required the child to, *"Make it so that there are more apples on this tree* (lesser number) *than on this tree"*, and similarly with appropriate reversal, for "less", children, once again, added to the tree indicated whether the request was to make it "more" or "less".

As this was a longitudinal study, the children were tested again six months later. The aims of this subsequent study was (a) to see if there was a change in response over the interval of time from the previous study to the present one, and (b) to verify the possibility that, responses to the questions involving "less" may have been influenced by the fact that they preceded those involving "more".

In this next study only one tree with a number of apples already on it was used. The child was asked, *"Make it so that there are less apples on the tree."*. After the child responded, s/he was then asked, *"Does it have less apples on it now?"*. There was no reference to "more" on this study, except in the case of two children who had in their response; they then were asked, *"Make it so that there are more apples on the tree."*.

Results, for this subsequent study, demonstrated that in the great majority of the cases (41 out of 52 responses), children treated "less" like "more" and added more apples to the tree.

Donaldson and Balfour (1968) conclude from these results that the children commonly and with very little sign of hesitation interpret it as a synonym for "more". Hence, this means that the children were responding as if they know the word "less" and as if they know it refers to quantity. Children seem to not differentiate "less" from "more", with "more" as the consistently dominant interpretation for the not differentiated pair.

Two critiques can be made of Donaldson and Balfour's study. (1) Donaldson and Balfour's conclusions are based only on the number of correct and incorrect responses out of a total number of questions asked. No appropriate statistical test was reported. In view of this, it is not clear whether there were significant differences. (2) The possibility also exists that the children may not have really understood what was required of them for questions containing the comparative term "less".

Palermo (1973) also investigated children's understanding of the comparative terms "more" and "less". Children were 3- to 7-years-old. Palermo replicated Donaldson and Balfour's (1968) study for both discontinuous and continuous quantities. The wooden apple trees were substituted for a pitcher of water in the continuous quantity condition.

Results from Palermo's study, also replicated those of Donaldson and Balfour (1968) for both types of quantities. Only four of the 16 children in the younger group and six of the 16 in the older group knew the concept "less". In a subsequent study with older children, Palermo again observed that even by age seven, some children have not yet differentiated between the two terms: "more" and "less".

Hence, there is evidence in Palermo's study that most children do not understand the concept of "less" in either a discontinuous or continuous fashion. Furthermore, these children would treat "less" like "more". These results were evident regardless of type of question asked and type of quantity presented.

Palermo suggests that this "more/less" phenomenon is a genuine cognitive characteristic of children at this age. However, these results, like those of Donaldson and Balfour (1968) do not give any indication as to why it is that children who do not know the concept "less", treat it as "more".

There is evidence to the contrary. Weiner (1974), for example, designed an experiment with the intention of replicating Donaldson and Balfour's results. Weiner examined 2- and 3-year-old children's understanding of the concepts "more" and "less" for questions asked about rows of toys that differed in the number of toys arranged in one-to-one correspondence to the point of inequality.

Results demonstrated that "more" is understood before "less", as children performed at chance level in most of the experimental conditions involving "less" questions.

However, there was no support for the findings of Donaldson and Balfour that children treat "less" as "more", as children were treating "less" randomly. Weiner concludes from the results obtained that the concept "less" is treated randomly.

This shows that when dealing with the comparison between quantities, there is a clear difference in the understanding of one comparative term and not another. Hence, it is wise to look at both comparative terms. Experiment 2, in this thesis, aims at investigating this type of comparative term, "less", in light of the other type, "more".

3.2 Method

(a) Subjects

Sixty-eight children, ranging in age from 5;1 - 6;3 (mean age of 5;7), divided into four groups of 17 children each served as subjects. Each group

of 17 children received one of the four conditions. One group of 17 children ranging in age from 5;5 - 6;3 (mean age of 5;8), received the discontinuous, comparative term "more" condition; another group of 17 children, ranging in age from 5;1 - 6;2 (mean age of 5;6), received the discontinuous, comparative term "less" condition; another group of 17 children, ranging in age from 5;3 - 6;3 (mean age of 5;9), received the continuous, comparative term "more" condition; and the final group of 17 children, ranging in age from 5;1 - 6;3 (mean age of 5;6), received the continuous, comparative term "less" condition. The children were all attending a First School in Oxford, England.

(b) Apparatus

A BBC Master microcomputer with twin 5-inch floppy diskettes was used to generate the charts, which were displayed on a 12-inch video monitor (Rediffusion computers, system alpha, model no. 364110). The program, written in BASIC, controlled the sequence of the displays, drew the static features of each display in high resolution graphics, and transferred information to and from data files stored on disc.

(c) Design

A mixed design with repeated measures on Conditions (4: Material [M], Comparative Term [CT], Question Type [Q], Order [O]) was used. Groups (Four groups of 17 children each: (1) Discontinuous material, comparative term "more", (2) Discontinuous material, comparative term "less", (3) Continuous material, comparative term "more", and (4) Continuous material, comparative term "less") was the Between factor.

Each child was given two sequences, in one session, in which there were 16 trials on each sequence. One sequence of 16 trials was for the

Equalize-type questions and the other sequence of 16 trials was for the Compare-type questions. Each of the 16 trials had a different presentation. There were 16 sequences of the 16 displays. Each child received one sequence for each question-type and the 17th child got a random sequence. Each different display had a trial label, represented by a number. Each sequence consisted of a randomized order of trials. These were the following:

1st sequence: 15-2-5-12-7-3-6-10-8-14-1-13-9-11-4-16
2nd sequence: 4-14-10-11-15-2-1-12-6-5-9-3-13-16-8-7
3rd sequence: 14-4-8-7-10-6-5-16-2-13-15-11-9-1-12-3
4th sequence: 15-14-1-10-8-4-13-11-12-6-3-7-5-9-2-16
5th sequence: 5-1-6-12-9-2-14-16-13-15-7-8-3-4-10-11
6th sequence: 1-13-5-14-11-10-4-15-12-3-9-16-6-7-8-2
7th sequence: 4-12-16-6-7-13-11-14-8-1-5-2-3-15-9-10
8th sequence: 15-12-13-4-3-10-14-8-9-7-6-1-16-11-5-2
9th sequence: 11-1-14-8-12-10-7-13-15-6-4-9-2-16-5-3
10th sequence: 13-14-9-4-6-11-12-3-2-8-7-15-1-10-16-5
11th sequence: 3-15-8-4-11-7-14-6-9-5-13-16-12-1-10-2
12th sequence: 2-6-8-10-13-11-5-16-7-9-14-3-1-12-15-4
13th sequence: 13-9-11-2-5-1-10-7-6-4-15-3-8-12-14-16
14th sequence: 5-11-4-13-7-6-10-16-8-1-9-15-2-12-14-3
15th sequence: 13-14-5-16-6-7-11-12-10-8-4-15-9-3-1-2
16th sequence: 11-13-8-5-4-14-3-7-9-6-15-12-10-16-2-1

The question types were presented in alternating order. Half of the children in one condition received the Equalize-type question first followed by the Compare-type question (sequence-set X); the other half were given the task in the reverse order (sequence-set Y). The design for Experiment 2 is illustrated in Table 3.2.1.

(d) **Materials**

Discontinuous Displays

Comparative Pair

On the screen, seven stimuli were presented. Two of these stimuli, which appeared on the left-hand side of the display, and which were separated from the other five stimuli, consisted of two vertical and parallel columns of discontinuous small circles. These two stimuli were the comparative stimuli, as the child had to determine the difference between them. They will be henceforth referred to as the "comparative pair". (See Figure 2.2.2.)

This comparative pair was taken randomly from a set of 8 small circles which were 1,2,3,4,5,6,7, and 8 inches high. The left-hand column of small circles was always the taller, labelled as the larger standard. It was either 8,7,6, and 5 inches high. The column of small circles next to it was always the shorter, labelled as the smaller standard. It varied from 1 to 7 inches in height in a Latin Square design. In one display, the smaller standard was 7 inches in height; in two, it was 6 inches in height; in three, it was 5 inches in height; in four, it was 4 inches in height; in three it was 3 inches in height; in two, it was 2 inches in height, and in one it was 1 inch in height. The difference in height between the smaller standard and the larger standard was always either 1,2,3, or 4 inches.

Hence, the design of the comparative pairs was as follows:

(larger standard) inches in height	1	2	3	4 (difference in height)
8	7A	6D	5D	4A
7	6D	5A	4A	3D
6	5A	4D	3A	2D
5	4D	3A	2D	1A
	(smaller standard) inches in height			

*A=ascending display in the right-hand set of items
*D=descending display in the right-hand set of items

Choice Stimuli

The remaining five stimuli, which appeared on the right-hand side of the display, consisted of five vertical and parallel columns of the same small circles. These five stimuli were the choice stimuli, as the child had to choose the one which was representative of the difference between the comparative pair. They will be henceforth referred to as the "choice stimuli". (See Figure 2.2.2.)

These choice stimuli were taken from a set of vertical columns that were 1,2,3,4,5,6,7,8, and 9 inches high. In half the displays, these columns of small circles on the right-hand side were arranged in ascending order and in the other half they were arranged in descending order. The set of items on the left-hand side was always a subset of the set of items on the right-hand side.

As mentioned in the design section, there were 16 fixed trials. The choice stimuli that was arranged in ascending order was as follows:

Comparative Pair	Choice Stimuli
8-7 inches in height	1,3,5,7, and 8 inches in height
8-4 inches in height	3,4,5,6, and 8 inches in height
7-5 inches in height	2,3,5,6, and 7 inches in height
7-4 inches in height	1,3,4,6, and 7 inches in height
6-5 inches in height	1,3,5,6, and 7 inches in height
6-3 inches in height	1,3,5,6, and 8 inches in height
5-3 inches in height	2,3,5,7, and 9 inches in height
5-1 inches in height	1,2,4,5, and 7 inches in height

The choice stimuli that was arranged in descending order was as follows:

Comparative Pair	Choice Stimuli
8-6 inches in height	8,6,5,4, and 2 inches in height
8-5 inches in height	8,6,5,3, and 1 inches in height
7-6 inches in height	7,6,5,3, and 1 inches in height
7-3 inches in height	7,5,4,3, and 2 inches in height
6-4 inches in height	7,6,4,2, and 1 inches in height
6-2 inches in height	6,5,4,3, and 2 inches in height
5-4 inches in height	5,4,3,2, and 1 inches in height
5-2 inches in height	7,6,5,3, and 2 inches in height

A Latin Square design was used to ensure that the occurrence of the correct response would appear an equal number of times in each of the five choice stimuli. Thus, the correct response was in the first position in four of the displays; in the second, third, fourth and fifth positions in three of them.

A total of 16 possible combinations was also devised so as to distribute the spaces equally between the choice stimuli in a Latin Square. With regard

to the five choice stimuli, the differences in height between adjacent stimuli were grouped in 16 possible combinations. These were as follows:

Difference number of inches in height between the choice stimuli

Number Distribution between each of the five choice stimuli in a Latin Square	Result of height of stimuli from Number Distribution
2 combinations of 1-1-1-1:	6,5,4,3,and 2 inches in height
	and,5,4,3,2,and 1 inches in height
1 combination of 1-2-1-2:	1,2,4,5,and 7 inches in height
1 combination of 2-1-2-1:	1,3,4,6,and 7 inches in height
1 combination of 1-2-2-1:	7,6,4,2,and 1 inches in height
1 combination of 2-1-1-2:	8,6,5,4,and 2 inches in height
1 combination of 1-1-1-2:	3,4,5,6,and 8 inches in height
1 combination of 2-2-2-1:	1,3,5,7,and 8 inches in height
1 combination of 1-1-2-2:	7,6,5,3,and 1 inches in height
1 combination of 2-2-1-1:	1,3,5,6,and 7 inches in height
1 combination of 2-1-1-1:	7,5,4,3,and 2 inches in height
1 combination of 1-2-2-2:	2,3,5,7,and 9 inches in height
1 combination of 1-2-1-1:	2,3,5,6,and 7 inches in height
1 combination of 2-1-2-2:	8,6,5,3,and 1 inches in height
1 combination of 1-1-2-1:	7,6,5,3,and 2 inches in height
1 combination of 2-2-1-2:	1,3,5,6,and 8 inches in height

In all displays, the items (rows of circles) in each set were positioned one inch apart, and there was a distance of 2 inches between the two sets of stimuli.

The colour of any given display was selected randomly. The displays could be yellow, green, blue, or pink in colour.

Continuous Displays

The continuous displays were similar to the discontinuous, except that the numerical quantities were here represented by bars. The same design was used as for the discontinuous material, except that, for example, 7 small

circles were replaced by a 7-inch bar. The number of small circles in the discontinuous condition always corresponded to the height in inches of the bars in the continuous condition. (See Figure 2.2.1 for Experiment 1.)

The differences between Experiment 1 and Experiment 2 can be summarized by the following points:

(1) Continuous and Discontinuous conditions were subdivided on the basis of the use of the comparative term "more" or "less". In Experiment 1, children were always asked, for example, *"How much more chocolate does the girl have than the boy?"*. Now they were also asked, *"How much less chocolate does the boy have than the girl?"*.

(2) Another difference was that the left-hand bar in the comparative pair was always the tallest, but not the same height across all the displays.

(3) Another difference was that the choice stimuli were arranged in descending as well as ascending order.

(4) The correct choice would occur equally often in each of the five choice stimuli.

(5) Finally, the number of displays in Experiment 2 was quadrupled from four illustrations in Experiment 1 to 16 illustrations.

(e) Procedure

Children were taken randomly from their classes individually. They were shown each display one at a time as displayed on the screen. Each of the 16 illustrations was displayed twice under the condition: (1) Question Type ("Equalize" or "Compare").

For each display, a child was asked two questions (16 questions in the Equalize form and 16 questions in the Compare form). Hence, each child was asked a total of 32 questions in all.

In all the displays, the larger standard of the comparative pair was the taller. For each display in the discontinuous condition, the experimenter asked one Compare-type question of the form, *"How many more does this one* (pointing to the larger standard) *have than this one* (pointing to the smaller standard)*?"*; and one Equalize-type question of the form, *"How many more does this one* (pointing to the smaller standard) *need to have, to have the same as this one* (pointing to the larger standard)*?"*. For each display in the continuous condition, the experimenter asked one Compare-type question of the form, *"How much more does this one* (pointing to the larger standard) *have than this one* (pointing to the smaller standard)*?"*; and one Equalize-type question of the form, *"How much more does this one* (pointing to the smaller standard) *need to have, to have the same as this one* (pointing to the larger standard)*?"*. The child then selected an answer (one of the five stimuli from the choice stimuli).

In the other comparative term version of the task, a child was asked again two questions for each display in the discontinuous condition: one of the Compare-type form, *"How many less does this one* (pointing to the smaller standard) *have than this one* (pointing to the larger standard)*?"*; and one of the Equalize-type form, *"How many less does this one* (pointing to the larger standard) *need to have, to have the same as this one* (pointing to the smaller standard)*?"*. For each display in the continuous condition, the experimenter would ask one Compare-type question of the form, *"How much less does this one* (pointing to the smaller standard) *have than this one* (pointing to the larger standard)*?"*; and one Equalize-type question of the form, *"How much less does this one* (pointing to the larger standard) *need to have, to have the same as this one* (pointing to the smaller standard)*?"*. The child then selected an answer (one of the five stimuli from the choice stimuli).

All the responses were recorded and stored in the computer.

3.3 Results

The following section will be divided into three sub-sections. One sub-section will be based on the pattern of correct and incorrect responses; the other sub-section will be based on the different observed strategies that children used to arrive at their responses; and the final sub-section will be based on the pattern of errors.

3.3.1 Comparing Correct Scores across Conditions

The principal data were the number of Equalize questions answered correctly, and the number of Compare questions answered correctly as in Experiment 1 (2.3). The mean number of correct responses for each type of question is shown in Table 3.3.1. These means were obtained from scores on the 16 questions on each question type and give an overall picture of children's performance. In general, once again, more correct responses were given to the Equalize questions than to the Compare questions with the exception of continuous "less", where the children obtained a very low score.

Table 3.3.1 Means and Standard Deviations (out of 16) for the Number of Correct Responses

The means and standard deviations for the total number of correct responses made in each condition by the children are given in Table 3.3.1. This table shows that Equalize questions were superior to Compare questions with discontinuous material and with continuous material when they were presented with the comparative term "more". In particular, Equalize questions were twice as good as Compare questions when presented with discontinuous material in the "more" comparative term. Overall, there was a difference between Equalize and Compare questions, as most correct responses were given to the former and not the latter. However, in one of the

cells, there was no difference in correct responses between Equalize- and Compare-type questions, as the children obtained such a low score. This was when these were presented with continuous material in the "less" comparative term.

Calculating the Probability of Correct Responses occurring purely by Chance

As children had a choice of five responses (from the choice stimuli), there was a 20% chance of their choosing any one of these responses and a mean chance score of 3.2. The probability that the correct responses occurred significantly above chance was explored using a modification of the standard hypothesis testing procedure with the means in Table 3.3.1.[3]

In this method, the observed proportions of responses (1, 2, 3, 4, 5) are compared with the predicted proportions obtainable by chance (.20) using the formula:

$$Z = \frac{\text{Observed Mean} - n.p}{\sqrt{n.p.q/N}}$$

where:

n = number of sequences
p = probability of any response occurring
q= (1 - p)
N = Number of subjects in Observed Mean

If a Z-value is greater than 1.96, the probability of the responses observed occurring by chance is less than 0.05. If a Z-value is greater than 2.58, the probability of the response occurring by chance is less than 0.01 (all values are for a 2-tailed test).

[3] The formula was personally communicated to this author by Dr. FHC Marriott, Statistics Department, University of Oxford.

Scores significantly above chance

The proportions of correct responses made in all, but two cell means, were significantly ($p<0.01$) greater than expected purely by chance. This indicates that the children were not acting randomly. The children were able as a group to solve the problems and pursue the right strategies.

The proportion of correct responses made by the group to the Compare question with continuous material, comparative term "more", was significantly ($p<0.05$) greater than expected purely by chance. This indicates that the children were not acting randomly and were pursuing the correct strategies.

Scores at chance

The proportion of correct responses made by the group to the Equalize question with continuous material, comparative term "less", was not significantly different from chance. This indicates that for this question, children were acting randomly.

3.3.2 Analysis of Variance

These trends were analyzed by subjecting the raw scores for the total number of correct responses in a 2x2x2x2 ANOVA. The main terms were Material (discontinuous or continuous), Comparative Term (more or less), Order (Equalize questions first, Compare second or Compare questions first, Equalize second), and Question Type (Equalize or Compare) with repeated measures on the last factor. The results of this analysis are presented in Table 3.3.2.

The significant main effect of Question ($F(1,60)=20.46$, $p<0.001$) indicated that Equalize questions once again were easier than Compare questions.

The significant interaction between Question Type and Comparative Term ($F(1,60)=7.64$, $p=0.01$) showed that the effects of the Comparative Terms affect significantly the children's performance on Question Type. This interaction was explored using a Newman-Keuls Multiple Range Test (Bruning and Kintz,1977:119-122). The Equalize and Compare difference is significant with the comparative term "more", but not with the comparative term "less". The use of the comparative term "more" produces a significantly higher number of correct responses from the children on the Equalize-type questions than on the Compare-type questions. Children are helped significantly by the use of "more" in Equalize-type problems, but not in the Compare-type problems. Children are just as bad on Equalize-type questions as on Compare-type questions when these are presented with the comparative term "less". Equalize questions are not helped with "less", but are helped with "more". A graphical representation of this comparison is presented in Figure 3.3.1.

The significant interaction between Question Type and Material ($F(1,60)=4.55$, $p<0.05$) showed that the effects of Material affects significantly the children's performance on Question Type. This interaction was explored using a Newman-Keuls Multiple Range Test. It was found that children perform significantly better on Equalize questions than Compare questions when discontinuous material is used. Children are helped significantly by the use of discontinuous material in Equalize problems, but not in Compare problems. Furthermore, Equalize questions with discontinuous material is significantly different from Equalize questions with continuous material and from Compare questions with continuous material. The Equalize and Compare difference is significant with discontinuous material, but not with continuous material. This result disagrees with the last experiment where there was no sign of a Question Type and Material interaction. A graphical representation of this comparison is presented in Figure 3.3.2.

3.3.3 Observing Children's Strategies

It is important to look at strategies as these give an indication of children's understanding of what is required, and, of their understanding of how they compare quantities and find out the difference between them. Hence, observations were recorded formally and informally on children's strategies in making their quantity judgements. For the 34 children performing the discontinuous condition, strategies were recorded informally. For the 34 children performing the continuous condition, strategies were recorded formally by video-taping them during their experimental session. It was found that there were four categories of responses, three of which were strategies recorded and based on the experimenter's observations of the children's performance on the task.

The Counting Strategy entailed counting the difference between the two comparative pairs, in the display presented, to arrive at a solution.

It is worth noting that sometimes, though very rarely, a one-to-one correspondence strategy (either in an observed cancelling out-type strategy or in an observed looking at the remainder-type strategy; the looking at the remainder-type strategy was observed when the child physically used a "cut-off point" between the two quantities in order to see what was left) was observed in conjunction with the counting strategy. For the purposes of this study, any indication of this simultaneous dual strategy was labelled as pertaining to the counting strategy.

The Measuring Strategy entailed an approximate measurement of the empty space from the smaller addend to the larger addend, which designated the difference between the two comparative pair, to arrive at a solution.

The Absolute Height Equivalence Strategy entailed choosing, of the choice stimuli, the equivalent in height of the larger standard or choosing the equivalent in height of the smaller standard to arrive at a solution. There were two types of Height Equivalence Strategy. One was the Height

Equivalence by Counting which entailed choosing the equivalent of one of the standards by counting; the other was the Height Equivalence by Height which entailed choosing the equivalent of one of the standards on the basis of height, only.

No Apparent Strategy was also accounted for. This entailed apparently choosing at random.

The Counting Strategy and the Measuring Strategy are correct strategies, whereas the Absolute Height Equivalence Strategy and the No Apparent Strategy lead to errors.

Percentage Number of Responses for each Strategy Type was recorded for both Equalize and Compare questions.

Strategy Types	Equalize Questions	Compare Questions
Counting Strategy	26%	19%
Measuring Strategy	18%	12%
Equivalence Strategy	51%	56%
No Apparent Strategy	5%	13%

It was found that for both Equalize and Compare questions, children preferred to use the Height Equivalence Strategy the great majority of the time.

Percentage Number of Responses for each Strategy Type was recorded for both discontinuous and continuous material.

Strategy Types	Discontinuous Material	Continuous Material
Counting Strategy	37%	10%
Measuring Strategy	0%	29%
Equivalence Strategy	55%	51%
Equivalence/Counting	40%	9%
Equivalence/Height	15%	42%
No Apparent Strategy	8%	10%

It was found that for both discontinuous and continuous material, children preferred to use the Height Equivalence Strategy, as found previously for the question types. There was a further finding in this type of condition. Children preferred to use Equivalence by Counting for the discontinuous material, whereas for the continuous material, they preferred to use an Equivalence by Height. It is worth noting that the Counting Strategy was widely used for the discontinuous quantity, whereas the Measuring Strategy was widely used for the continuous material.

Percentage Number of Responses for each Strategy Type was recorded for both comparative terms: "more" and "less".

Strategy Types	More Comparative Term	Less Comparative Term
Counting Strategy	30%	16%
Measuring Strategy	22%	8%
Equivalence Strategy	37%	69%
Equivalence/Counting	22%	27%
Equivalence/Height	15%	42%
No Apparent Strategy	11%	7%

It was found that for both comparative terms, children preferred to use the Height Equivalence Strategy, as found previously for the question types and for material types. Children tended to choose the equivalent of the larger standard for the "more"-type questions, and/or the equivalent of the smaller standard for the "less"-type questions to arrive at a solution. There was, also, a further finding in this type of condition. Children preferred to use Equivalence by Counting for the comparative term "more" ("counting-up" was observed for the "more" comparative term [i.e. for addition] and "counting-down" was observed with the "less" comparative term [i.e. for subtraction]), whereas for the comparative term "less", they preferred to use an

Equivalence by Height. Another interesting observation is that for the comparative term "more", children demonstrate a high percentage use of the counting strategy indicating that this variable is easier understood than any of the other variables.

At this stage it is interesting to note that in this Equivalence Strategy, children did not just choose the equivalent of the larger standard in the "more"-type question and the equivalent of the smaller standard in the "less"-type question, but could also choose the smaller standard in the Equalize-"more" question and the larger standard in the Equalize-"less" question. This distinction was only observed for the Equalize-type questions only.

It is concluded that children showed a tendency to use the correct strategies with Equalize problems more so than with Compare problems.

3.3.4 Children's Errors

It is also important to look at children's errors as these also give an indication of children's understanding of how they compare quantities and find out the difference between them. An errors' analysis was performed by looking in more detail at children's individual responses to each question type under each comparative term. There were four categories of errors, which were recorded, based on the child's choice stimuli when performing the task.

The Same Comparative Pair Larger Standard Error [SL] entailed choosing the choice stimuli that was the same as the larger standard in the comparative pair, in the display presented, to arrive at a solution.

The Same Comparative Pair Smaller Standard Error [SS] entailed choosing the choice stimuli that was the same as the smaller standard in the comparative pair, in the display presented, to arrive at a solution.

The Extreme Larger Error [EL] entailed choosing the largest of the choice stimuli, in the display presented, to arrive at a solution.

The Extreme Smaller Error [ES] entailed choosing the smallest of the choice stimuli, in the display presented, to arrive at a solution.

Other Error was also accounted for. This entailed choosing one of the remaining two choice stimuli that were not either the [SL], [SS], or [EL and ES].

The Total Number of Errors for each error type was recorded for both Equalize and Compare questions with both comparative terms: "more" and "less".

Error Types	**Equalize More**	**Compare More**	**Equalize Less**	**Compare Less**
Total Errors	238	378	339	381
Same Larger Standard	088	245	065	070
Same Smaller Standard	034	028	184	230
Extreme Largest	080	199	050	060
Extreme Smallest	079	046	089	090
Other	028	035	039	034

It was found that children made the most amount of errors on the Compare questions than on the Equalize questions. Upon a further breakdown of the errors, it was found that children made these errors depending on which comparative term was presented and regardless of question type. Children made the Same Comparative Pair Larger Standard Error on the "more"

comparative term-type questions, whereas they made the Same Comparative Pair Smaller Standard Error on the "less" comparative term-type questions. Children also made the Extreme Larger Error on the "more"-type questions, and the Extreme Smaller Error on the "less"-type questions.

3.4 Discussion

The results of Experiment 2 confirm that children do not perform well on the type of problem which asks them to compare quantities, but can solve Equalize-type word problems reasonably well, particularly if these are presented in discontinuous fashion and using the "more" comparative term. The Equalize and Compare difference result replicated that of Experiment 1. Hence, some generality can be made about the method of presenting the material. The result will not alter, whether the method of presentation involves the use of concrete material or of a computer.

Why are Equalize problems so difficult with "less"? One hypothesis involves the concept of reversibility. Children at the age of 5-to-6-years seem to have difficulty with this concept. Very few of the children could think of "less" in terms of "more". One child demonstrated understanding of this concept when she elicited, *"If you put, it'll be the same, so if you take away, it'll be the same too."*. This little girl was an exceptional case. Most of the children did not demonstrate understanding of this concept, hence their unsuccessful performance with the comparative term "less" in both type questions.

A second hypothesis involves the dynamic concept. Children have been demonstrated to be better on dynamic problems when these involve a change-up, rather than when these involve a change-down.

This experimenter considers that there is evidence that the reason why children have difficulty with Compare-type problems is not due to a linguistic

component. Different linguistic markers were used in Experiment 2. Linguistic markers affect performance on the Equalize-type problems in the sense that children do better with "more" and worse with "less". However, linguistic markers do not affect performance on the Compare-type problems and children still do poorly on this type of problem.

Whatever strategy the children use on Equalize word problems (either a one-to-one correspondence strategy, as has previously been suggested by Hudson, or a counting strategy, as was also further suggested by Hudson and by this experimenter, as well as a measuring strategy, also suggested by this experimenter), it is not sufficient to tackle the Compare word problem. Whatever strategy is needed to solve the Compare-type problem, children do not seem to possess it.

It seems that children in the continuous task use an equivalent strategy to the one of counting in the discontinuous task. This equivalent strategy is a measuring one. However, this seems to be true only for the Equalize-type questions in the "more" comparative term, which they understand.

For the Compare-type question, as well as for both type questions in the "less" comparative term, children use an Equivalence Strategy whereby they choose the response on the basis of counting for discontinuous material and on the basis of height for continuous material. They seem to use this strategy when they find the word problem too difficult to handle. This strategy does show that they do have some knowledge of "more" and "less", "bigger" and "smaller", "taller" and "shorter", as children choose the equivalent of the larger standard in the "more" Compare-type question and the equivalent of the smaller standard in the "less" Equalize- and Compare-type question. This was further supported by the analysis of errors, where children exhibited a strong tendency to choose the choice stimuli that was the same as one of the comparative pair, depending on the comparative term used. The Equivalence Strategy yielded these type of errors, along with the extreme-type errors,

which were also quite evident in the Compare-type questions and in the Equalize-type questions with "less".

Overall, it can be concluded that children seem to misunderstand the Compare-type question. Based on the analysis of errors, there are three hypotheses that may explain such a misunderstanding. One hypothesis suggests that when children are asked, *"How many more does this one have than this one?"*, they cannot choose the correct choice and instead choose the one that has more. As pointed out in Section 1.7.3, this is supported by Moore and Frye's (1986) and Bloom's (1970) claim, that children understand "more" in the sequential sense and do not acquire the simultaneous meaning until about the age of five. A second hypothesis suggests that children cannot understand that what they are being asked to do is to break up the whole that has more into two bits. This suggests that children's misunderstanding lies in their knowledge of part-whole relationships. The final hypothesis involves, as previously mentioned in this thesis, the dynamic concept. A Compare-type problem, requires the child to refer back to an initial state of the start-set in order to see what the difference between the two quantities actually is. (See Section 1.7.1.3.) As the child may be unable to do this, s/he remains in the initial state and as the analysis of errors proves, chooses that initial state, which is the start set, as the correct answer.

It seems that the concept of "less" is too difficult for children of 5- and 6-years-old to grasp as they cannot even do well in the Equalize version of the question with this comparative term. However, the fact that, when performing this strategy, children may also choose the equivalent of the smaller standard for the Equalize-type question with the "more" comparative term and the equivalent of the larger standard for the Equalize-type question with the "less" comparative term, demonstrates that children are closer at achieving the correct response for Equalize-type questions. They

demonstrate some understanding by attempting to cognitively decipher this distinction.

There may be a developmental sequence where these strategies are concerned. A cognitive model can be created. (See Figure 3.4.1.) Children start by not understanding what is going on. From this, children exhibit equivalence. Through this strategy the child demonstrates some sign of understanding quantitative relationships. If s/he did not have any understanding of these relationships, s/he would have demonstrated No Apparent Strategy by choosing at random or as in most cases observed, always choosing the largest item of the choice stimuli. When it is too difficult for the child to process the question, s/he exhibits equivalence. From this strategy, the child then develops the acquisition of either a Measuring Strategy used for continuous material, or, a Counting Strategy for discontinuous material. Thus, a three-step developmental sequence may be required to achieve an appropriate strategy for word problem solving.

However, this sequence is more noticeable for Equalize-type questions. For Compare-type questions, children seem to stumble on the Equivalence Strategy, which triggers exactly the same reaction when the comparative term, "less", is used.

The Equivalence Strategy is an interesting strategy to explore as it may be the transition ground the child must travel when s/he goes from No Apparent Strategy to the acquisition of a Measuring or Counting Strategy. This strategy suggests that the child has some understanding of the concept of absolute size, as well as a partial notion of "more" and "less", "greater" and "lesser", and part-whole relationships, as s/he understands about the whole. However, this strategy also indicates that the child still lacks an understanding about the parts that make up that whole. Experiment 3 was designed to look at children's understanding of part-whole relationships via equivalence.

FIGURES AND TABLES FOR EXPERIMENT 2

Table 3.2.1 The Design of Experiment 2

No. of children	Material Group	Comparative Term	No. of children	Question Type (Seq.Set)
17	DISCONTINUOUS	More	8	x
			9	y
17		Less	9	y
			8	x
17	CONTINUOUS	More	9	y
			8	x
17		Less	8	x
			9	y
N=68				

x = Sequence Set Equalize first, Compare second
y = Sequence Set Compare first, Equalize second

Table 3.3.1 Means and Standard Deviations (out of 16) for the Number of Correct Responses

Material Group	Comparative Term	Question Type	
		Equalize	Compare
Discontinuous	More	10.48 (4.36)	5.08 (5.74)
	Less	7.96 (4.73)	5.25 (4.48)
Continuous	More	7.91 (4.71)	4.43 (5.39)
	Less	4.27 (2.94)	4.84 (3.91)

(Standard deviations in brackets.)

Table 3.3.2 Summary Table (ANOVA) for Total Correct Scores

Source	SS	DF	MS	F	PROB.
Mean	5336.851	1	5336.851	173.00	0.0000
Material[M]	113.236	1	113.236	3.67	0.0601
ComparativeTerm[C]	65.851	1	65.851	2.13	0.1492
Order [O]	64.870	1	64.870	2.10	0.1522
MC	1.596	1	1.596	0.05	0.8209
MO	65.851	1	65.851	2.13	0.1492
CO	1.753	1	1.753	0.06	0.8124
MCO	3.883	1	3.883	0.13	0.7240
error	1850.938	60	30.849		
Question[Q]	257.856	1	257.856	20.46	0.0000
QM	57.292	1	57.292	4.55	0.0371
QC	96.287	1	96.287	7.64	0.0076
QO	2.582	1	2.582	0.20	0.6525
QMC	3.963	1	3.963	0.31	0.5770
QMO	10.920	1	10.920	0.87	0.3557
QCO	0.008	1	0.008	0.00	0.9796
QMCO	0.005	1	0.005	0.00	0.9842
error	756.132	60	12.602		

SS = Sum of Squared Deviations
DF = Degrees of Freedom
MS = Mean of Squared Deviations
F = Ratio of Variances
Prob. = Level of Significance

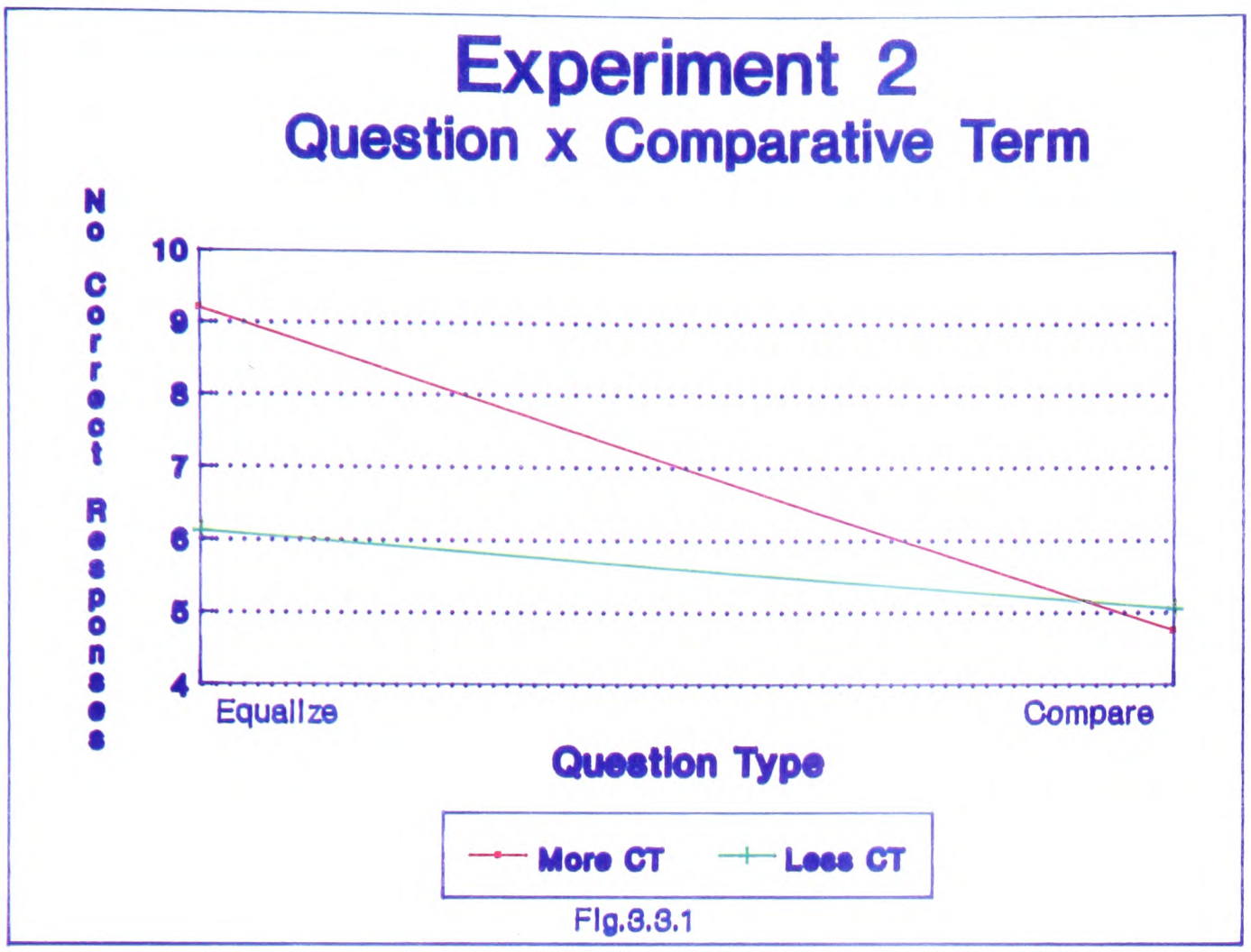


Fig.3.3.1

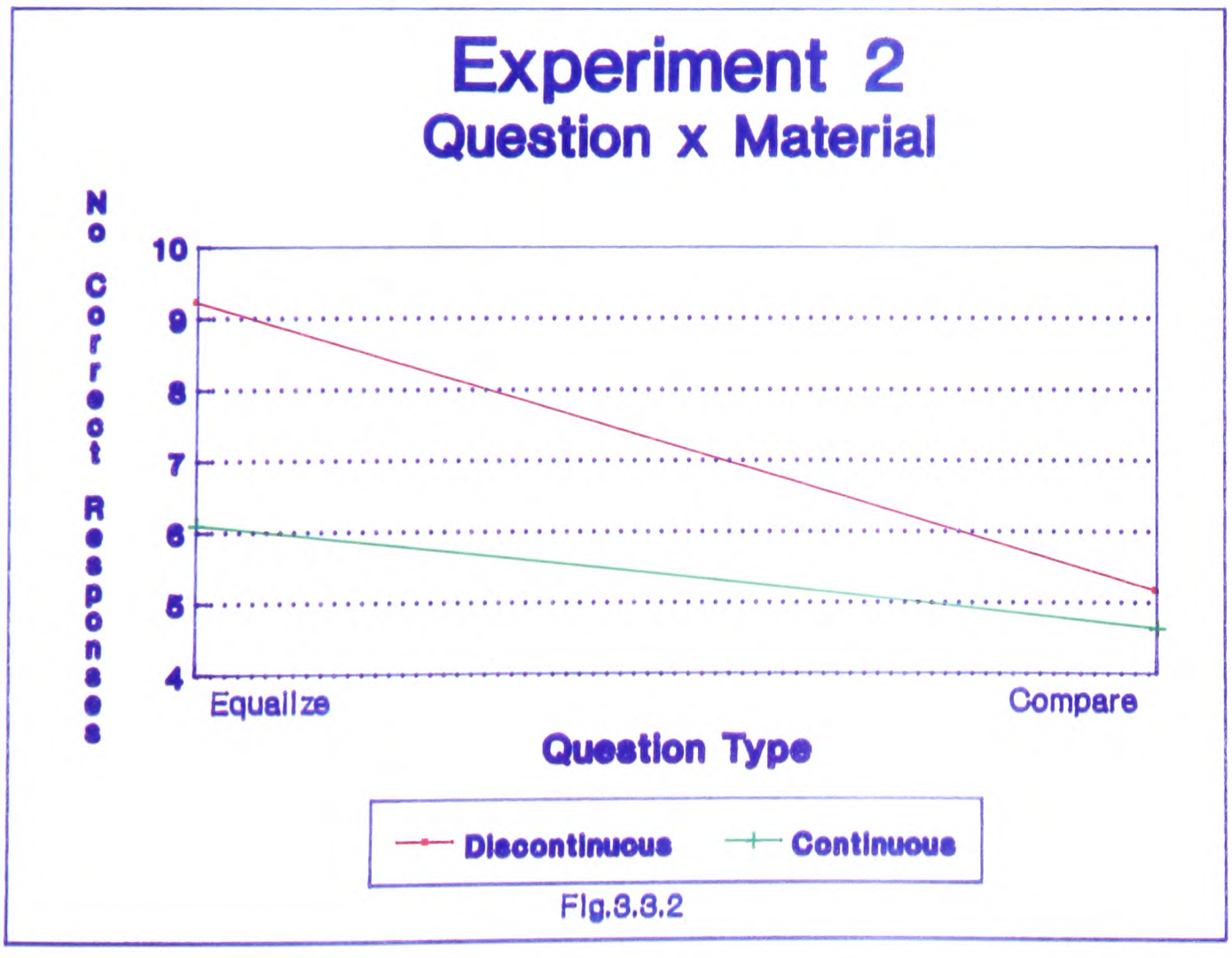
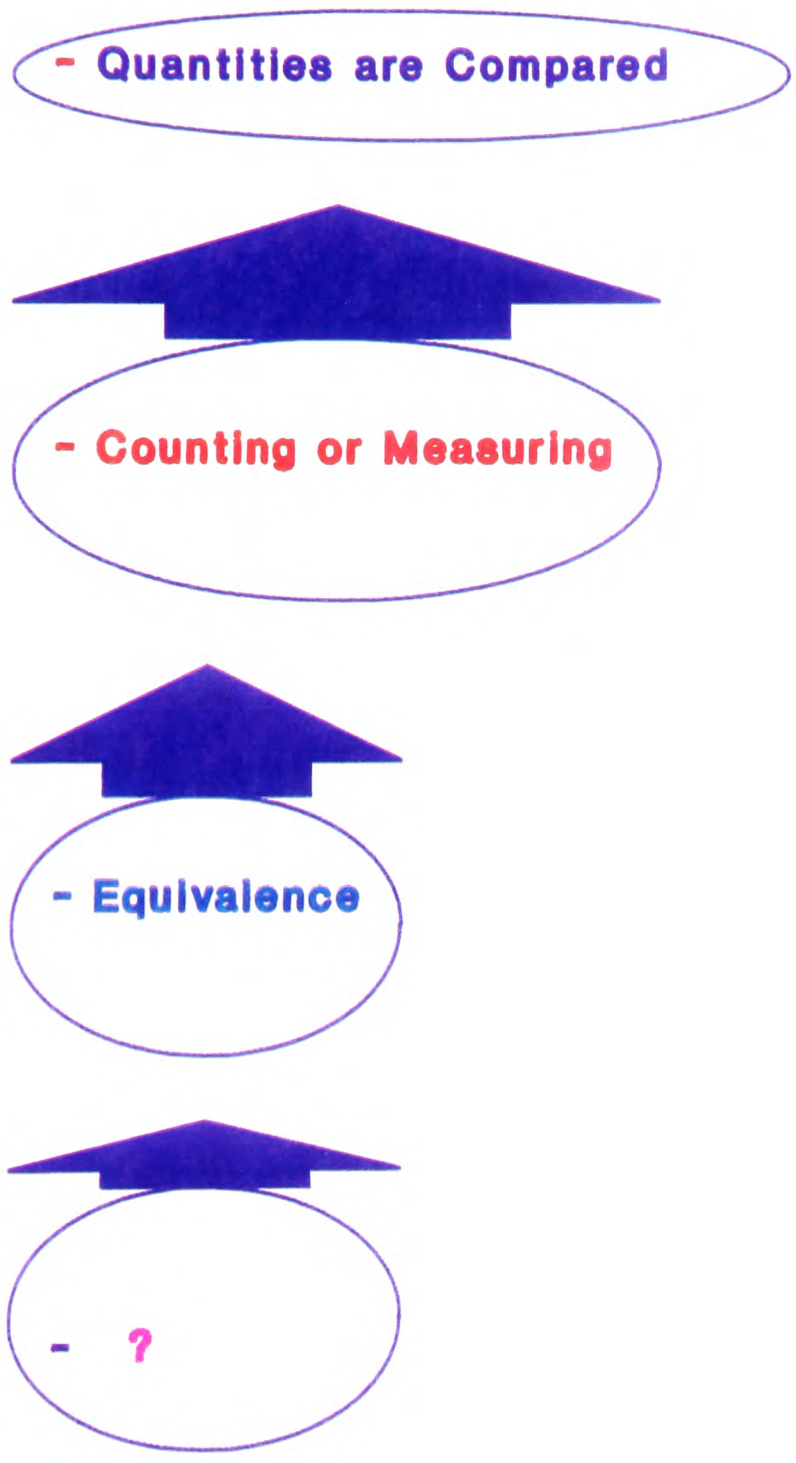


Fig.3.3.2

Table 3.4.1 Model for the Successful Performance of Quantitative Comparisons



CHAPTER FOUR

PART-WHOLE RELATIONSHIPS WITHIN QUANTITATIVE COMPARISON

4 Experiment 3 Part-Whole Relationships via Equivalence

4.1 Introduction

The results of the first two experiments are consistent in producing a difference between Equalize and Compare problems and show that young children (5-to-6-years old) find more difficulty with Compare than with Equalize word problems when they are presented with discontinuous and continuous graphical information, either on concrete material or on a computer screen.

An explanation is needed as to why static Compare questions are more difficult than action Equalize questions. One hypothesis, which is the basis for the present experiment (Experiment 3), deals with part-whole relationships. The reasoning behind this hypothesis is based on the speculation that, in Experiments 1 and 2, when children are given a static Compare-type question, they may need to understand that the larger comparative stimulus has two parts, one part which is equivalent to the smaller comparative stimuli and the other part which is the remainder or difference between the two stimuli. This may not be so for Equalize problems. When children are given an action Equalize-type question, they may understand it in another way. They may understand that the smaller comparative stimuli has to be increased to make it the same as the larger comparative stimuli. Children may then use that increase as the solution to

the problem. Hence, in contrast to Compare problems, children may avoid having to use the part-whole relationship in an Equalize problem.

The present experiment was then designed to investigate the effect of giving children strong perceptual cues about the part-whole relationship. It was thought that making the parts of the larger stimulus explicit would help children to perform more successfully on Compare word problems. Hence, lines were set up, horizontally, dividing the larger stimulus of the comparative pair for the continuous material. (See Figure 4.1.1.) In one condition, the line divided the larger bar into two separate portions, one of which was the same size as the smaller comparative stimulus, while the other represented the difference between the two; thus, demonstrating explicitly how the whole has two parts to it. It was predicted that this manipulation would increase performance on Compare problems more than on Equalize problems, since the hypothesis was that the understanding of part-whole relationships was not essential for the solution of Equalize problems.

In order to avoid the possibility of attributing children's performance to their merely dividing the larger standard, a control condition was included. In this control condition, the dividing line did not divide the larger standard of the comparative stimuli into a part equivalent to the smaller standard and a part equivalent to the difference between the two, but rather into a part which was not the same size as the smaller standard and another part which was not the same size as the remainder, which is the difference between the two comparative stimuli.

4.2 Method

(a) Subjects

Ninety-six children, ranging in age from 5;1-7;8 (mean age of 6;1), divided into six groups of 16 children each served as subjects. Each group of 16 children received one of six conditions. One group of 16 children, ranging in age from 5;3 - 6;3 (mean age of 5;8), received the continuous, comparative term "more" condition; another group of 16 children, ranging in age from 5;3 - 6;3 (mean age of 5;6), received the continuous, comparative term "less" condition; another group of 16 children, ranging in age from 5;9 - 7;7 (mean age of 7;1), received the "A1=B", comparative term "more" condition; another group of 16 children, ranging in age from 5;2 - 6;9 (mean age of 5;7), received the "A1=B", comparative term "less" condition; another group of 16 children, ranging in age from 5;2 - 7;8 (mean age of 5;7), received the "A1≠B", comparative term "more" condition; and the final group of 16 children, ranging in age from 5;2 - 7;7 (mean age of 6;4), received the "A1≠B", comparative term "less" condition. The children were all attending a first school in Oxford, England.

(b) Apparatus

The same apparatus was used as for Experiment 2.

(c) Design

There were 96 children, and they were divided into three groups: one given Undivided comparative stimuli, one given comparative stimuli in which the larger stimulus was divided into two sections, one of which equalled the smaller comparative stimulus (A1=B), and the third given comparative stimuli in which the larger stimulus was divided out into two sections, but neither section was equivalent to the smaller comparative stimulus (A1≠B). (See

Figures 2.2.1 and 4.1.1.) The comparative term used for half of each group was "more", and for the other half "less". All children were given both questions (Equalize and Compare), and half of each sub-group was given Equalize before Compare and the other half Compare before Equalize. Therefore, the design was 3 [Material] x 2 [Comparative Term] x 2 [Order] x 2 [Question Type] with repeated measures on the last factor.

Each child was given two sequences, in one 32-trial session, in which there were 16 trials for each sequence. One sequence of 16 trials was for the Equalize-type questions and the other sequence of 16 trials was for the Compare-type questions. Each of the 16 trials had a different presentation. There were 16 sequences of the 16 displays. The same sequences of 16 displays were used for all three groups: Undivided displays, "A1=B" displays, and "A1≠B" displays. Each child received one sequence for each question type. Each different display had a trial label, represented by a number. Each sequence consisted of a randomized order of trials. These were the following:

1st sequence:	15-9-5-1-11-7-4-12-8-6-3-2-14-13-10-16
2nd sequence:	1-13-9-12-6-3-11-14-16-10-15-7-2-5-4-8
3rd sequence:	2-10-15-16-12-11-6-14-5-7-1-4-9-13-8-3
4th sequence:	10-16-8-13-12-5-4-1-15-14-9-2-6-11-3-7
5th sequence:	10-6-14-16-2-3-15-9-5-7-11-4-1-12-8-13
6th sequence:	10-8-15-3-9-6-14-12-1-11-2-7-16-5-4-13
7th sequence:	13-16-15-14-7-12-6-2-11-4-8-5-10-3-1-9
8th sequence:	4-7-8-16-15-5-10-9-6-3-1-12-13-14-11-2
9th sequence:	14-11-9-2-3-10-6-1-4-12-16-5-13-7-8-15
10th sequence:	14-3-4-16-1-7-13-5-15-6-2-12-10-9-11-8
11th sequence:	12-6-13-16-10-2-7-5-8-4-9-11-14-1-3-15
12th sequence:	8-15-9-16-2-7-4-12-6-10-3-5-13-14-11-1
13th sequence:	16-4-8-5-10-3-15-11-7-14-1-6-13-9-12-2
14th sequence:	10-5-4-3-9-8-11-7-16-1-2-12-6-15-14-13
15th sequence:	2-7-4-16-10-9-12-11-1-8-15-3-5-13-6-14
16th sequence:	9-3-15-10-6-8-7-12-1-4-11-14-13-16-5-2

The question types were presented in alternating order. Half of the children in one group received the Equalize-type question first followed by the Compare-type question (sequence-set X); the other half were given the task in the reverse order (sequence-set Y). The design for Experiment 3 is illustrated in Table 4.2.1.

(d) Material

Each of the three groups was given a different type of display. The displays, for all three groups, were the same in size as for Experiment 2. [See Section 3.2 (d).] The different displays were the following:

Undivided Displays

The Undivided displays were the same continuous computer-driven displays used in Experiment 2. (See Figure 2.2.1.)

"A1=B" Displays

An additional element in the displays was a line drawn across the larger standard of the comparative stimuli, corresponding to the height of the smaller standard (known here as "B"). The part of the larger standard which was equal to "B" was called "A1", and the remaining part was called "A2". In this condition "A2" was always the correct answer (i.e. the difference between the two standards), and found among the choice stimuli. (See Figure 4.1.1.)

"A1≠B" Displays

In this condition, the addition on the display was again a line drawn across the larger standard of the comparative stimuli, but not corresponding to the height of the smaller standard (known here as "B"). This line was always drawn across the larger standard of the comparative stimuli, upon the

calculation of A1=B+1 inch. Again, here, there were two parts to the larger standard, "A1", which was the bottom-half part and "A2" which was the top-half part. In this condition "A2" was never the correct answer to be found among the choice stimuli. (See Figure 4.1.1.)

(e) **Procedure**

The same procedure was used on all three conditions as in Experiment 2 for the continuous condition.

4.3 Results

The following section will be divided into two sub-sections. One sub-section will be based on the pattern of correct and incorrect responses; the other sub-section will be based on the different observed strategies that children used to arrive at their responses.

4.3.1 Comparing Correct Scores across Conditions

The principal data were the number of Equalize questions answered correctly, and the number of Compare questions answered correctly as in Experiments 1 (2.3) and 2 (3.3). The mean number of correct responses for each type of question is shown in Table 4.3.1. These means were obtained from scores on the 16 questions on each question type and give an overall picture of children's performance.

It was predicted that children would perform better in the "A1=B" displays for the Compare questions, but not for the Equalize questions. A general improvement in the Compare questions with the "A1=B" displays was expected.

In general, once again, more correct responses were given to the Equalize questions than to the Compare questions, and the prediction was not supported in the "A1=B" displays. It so seems that adding a line in the comparative stimuli makes it worse for Compare questions in both types of displays ("A1=B" and "A1≠B").

Table 4.3.1 **Means and Standard Deviations (out of 16) for the Number of Correct Responses**

This table shows that for the Undivided displays, the Equalize and Compare difference is not confirmed with the comparative term "more" for the continuous displays, and is hardly confirmed with the comparative term "less" for the same-type displays. This is an inconsistency that can not be explained.

The table also shows that children's performance did not improve on the Compare questions, as compared to the Undivided displays, for the "A1=B" displays. Performance remained about the same on the Equalize questions, as compared to the Undivided displays, for the "A1=B" displays.

These are puzzling results, as it was predicted that children's performance would increase in the Compare problems in this display-type, as the line dividing the whole into two parts (one of which was clearly the difference between the two quantities) was there as an aid. However, the above is valid for the comparative term "more" and children did worse with "more" in this display type, but showed an improvement between the "A1=B" displays and the Undivided displays with "less". There was an even greater improvement with "less" in the "A1≠B" displays. Perhaps children are able to concentrate better on "less" when this dividing line is present. Children's "less" comparative judgments are improved with this dividing line. This was not predicted. Furthermore, the same pattern is observed with the Equalize

questions. It thus seems that dividing the larger comparative stimulus improves "less" judgements on Equalize questions, as well.

This table also shows that for the "A1≠B" displays, roughly the same pattern of responses was made in the Equalize questions and in the Compare questions. There was a striking superiority in the Equalize questions and in the Compare questions in this display-type over the Undivided displays and the "A1=B" displays. These negative results were inconsistent with previous results. The Equalize and Compare difference was not replicated in the Undivided displays. However, the Equalize and Compare difference was significant in the basic pattern of results in the "A1=B"- and "A1≠B"-type of displays where dividing the comparative stimulus made "more" judgements harder and "less" judgements easier.

Calculating the Probability of Correct Responses occurring purely by Chance

As for Experiment 2, children had a choice of five responses (from the choice stimuli). Therefore, there was a 20% chance of their choosing any one of these responses. Children were considered to be performing at chance level if they got a total of 3.2 correct responses. The probability that the correct responses occurred simply by chance was explored using a modification of the standard hypothesis testing procedure with the means in Table 4.3.1. (See Section 3.3.1.)

From these calculations, there were three kinds of results whereby (1) some results might be significantly below chance, (2) some at chance, and (3) some above chance.

Scores significantly below chance

The proportion of correct responses made by two groups: (1) Equalize question with Undivided material, comparative term "less", and (2) Compare question with Undivided material, comparative term "less", were significantly ($p<0.05$ and $p<0.01$, respectively) below chance level. This indicates that for these two groups, children were not acting randomly, but systematically adopted a wrong strategy. Though the children in these two groups were avoiding the correct stimulus, and were significantly worse than the children in the other groups, (see Table 4.3.1), they had a definite hypothesis when performing the task. This hypothesis, however, was the wrong one. Children's hypotheses will be examined in the next section which describes their observed strategies.

Scores at chance

The proportion of correct responses made by two groups: (1) Compare question with "A1=B" material, comparative term "less", and (2) Compare question with "A1≠B" material, comparative term "more", were not significantly different from chance. This indicates that for these two groups, children were acting randomly.

Scores significantly above chance

The proportion of correct responses made to the Compare question with "A1=B" material, comparative term "more", was significantly ($p<0.05$) greater than expected purely by chance. The proportions of correct responses made in the remaining cell means (see Table 4.3.1) were significantly ($p<0.01$) greater than expected purely by chance. This indicates that for these groups, the children were not acting randomly and were pursuing the correct strategies.

4.3.2 Analysis of Variance

These trends were analyzed by subjecting the raw scores for the total number of correct responses to a 3x2x2x2 ANOVA in which the main terms were: Material (Undivided, "A1=B", "A1≠B"), Comparative Term (more or less), Order (Equalize questions first, Compare second versus Compare questions first, Equalize second), and Question Type (Equalize or Compare), with repeated measures on the last factor. The results of this analysis are presented in Table 4.3.2.

The hypothesis that children would perform better in the "A1=B" displays for the Compare questions, but not for the Equalize questions predicts a Material x Question Type interaction. This interaction, however, did not occur.

As discussed earlier, there was a difference in performance on "more" and "less" in the different groups. There was an improvement with "less" in the "A1=B" group, as well as in the "A1≠B" group, and there was a decrease in performance with "more" in the "A1=B" group, as well as in the "A1≠B" group. The divided conditions seemed to be better for "less", but worse for "more". Therefore, a Material x Comparative Term interaction was expected. However, this interaction did not occur either.

The significant main effect of Question ($F(1,84)=19.09$, $p<0.001$) indicated that Equalize questions were once again easier than Compare questions.

The significant interaction between Question Type and Comparative Term ($F(1,84)=5.12$, $p<0.05$) showed that the effect of the Comparative Terms affects significantly the children's performance on different Question Type. This interaction was explored using a Newman-Keuls Multiple Range Test (Bruning and Kintz,1977:119-122). It was found that the linguistic comparative marker is important. The use of the comparative term "more" on the Equalize-type questions produces a significantly higher number of correct

responses from the children than on the Compare-type questions. Children are helped significantly by the use of "more" in Equalize-type problems. The Equalize and Compare difference is only significant with the "more" comparative term, but not with the "less". This interaction replicated the one in Experiment 2. A graphical representation of the interaction is presented in Figure 4.3.1.

The significant interaction between Material and Order ($F(2,84)=6.93$, $p<0.01$) showed that the effect of order presentation of question type affects significantly the children's performance on Material. Children seem to perform significantly better on the "A1≠B" displays when the Compare-type questions are asked first, followed by the Equalize-type questions than on the Continuous displays when the Compare-type questions are asked first, followed by the Equalize-type questions. It seems that in this most difficult display, children are helped significantly by being presented with the Compare-type question first, followed by the Equalize-type question. This seems not to be the case in the other displays, as these were not as difficult. However, a Newman-Keuls Multiple Range Test applied to this interaction demonstrated that each condition did not differ significantly from each other. A graphical representation of this comparison is presented in Figure 4.3.2.

The significant three-way interaction between Question and Comparative Term and Order ($F(1,84)=4.89$, $p<0.05$) showed that the effect of order presentation of question type affects significantly the effect of Comparative Term on the children's performance on Question Type. This interaction was explored using a Newman-Keuls Multiple Range Test (Bruning and Kintz,1977:119-122). It was found that the children performed the worst when the Compare-type question in the "less" comparative term followed the Equalize-type question and when the Compare question in the "more" comparative term was presented first. When the Equalize question in the "less" comparative term was presented first, it was also significantly

different from the other variables. A graphical representation of this interaction is presented in Figure 4.3.3.

4.3.3 Observing Children's Strategies

Observations were again recorded on children's strategies in making their quantity judgments. The same four categories of responses were recorded as for Experiment 2. (See Section 3.3.2.)

Percentage Number of Responses for each Strategy Type was recorded for both Equalize and Compare questions.

Strategy Types	Equalize Questions	Compare Questions
Counting Strategy	7%	5.5%
Measuring Strategy	30%	18.5%
Equivalence Strategy	60%	60.5%
No Apparent Strategy	3%	15.5%

It was found, as in Experiment 2, that for both Equalize and Compare questions, children preferred to use the Height Equivalence Strategy most of the time. However, there was also a different pattern for the Equalize and Compare questions. No Strategy was more evident in Compare questions than in Equalize. It is also worth noting that the Measuring Strategy was widely used in the Equalize-type question and not so much so in the Compare-type question. The Measuring Strategy is a substitute for the Counting Strategy, for continuous material. Children seem to use it more frequently, as they use the Counting Strategy for discontinuous material, when they understand the question.

Percentage Number of Responses for each Strategy Type was recorded for the three display types: Continuous, "A1=B", and "A1≠B".

Strategy Types	Continuous Display	"A1=B" Display	"A1≠B" Display
Counting Strategy	10%	6%	2%
Measuring Strategy	28%	22%	23%
Equivalence Strategy	51%	64%	65%
Equivalence/Counting	10%	2%	0%
Equivalence/Height	41%	62%	65%
No Apparent Strategy	10%	8%	10%

It was found, as in Experiment 2, that for all three display types, children preferred to use the Height Equivalence Strategy, as found previously for the question types. There was a further finding in this type of condition. Children preferred to use Equivalence by Height for all three types of display. This replicated findings in Experiment 2 for the continuous material. It is worth noting that the Measuring Strategy was widely used in these displays.

Percentage Number of Responses for each Strategy Type was recorded for both comparative terms: "more" and "less".

Strategy Types	More Comparative Term	Less Comparative Term
Counting Strategy	10%	3%
Measuring Strategy	30%	18%
Equivalence Strategy	48%	72%
Equivalence/Counting	6%	1%
Equivalence/Height	42%	71%
No Apparent Strategy	12%	7%

It was found, as in Experiment 2, that for both comparative terms, children preferred to use the Height Equivalence Strategy, as found previously for the

question types and for display types. Children tended to choose the equivalent of the larger standard for the "more"-type questions, and/or the equivalent of the smaller standard for the "less"-type questions to arrive at a solution. The Measuring Strategy was widely used for the comparative term "more". This was also true for the Equalize-type questions. It is interesting that these two variables triggered widespread use of the Measuring Strategy. Furthermore, it is interesting to note that there was an interaction of Question Type and Comparative Term. It is speculated that this interaction triggered use of this strategy type.

Percentage Number of Responses for each Strategy Type was recorded for both Equalize and Compare questions with both comparative terms: "more" and "less".

Strategy Types	Equalize More	Compare More	Equalize Less	Compare Less
Counting Strategy	11%	9%	3%	2%
Measuring Strategy	40%	21%	20%	16%
Equivalence Strategy	45%	51%	74%	70%
Equivalence/Counting	8%	4%	1%	2%
Equivalence/Height	37%	47%	73%	68%
No Apparent Strategy	4%	19%	3%	12%

As found for percentage number of responses for each strategy type recorded for question type, children preferred to use the Height Equivalence Strategy the great majority of the time. The use of this strategy type is very high for the Equalize-type questions with the "less" comparative term (73%) and for the Compare-type questions with the "less" comparative term (68%).

This demonstrates that children had a definite hypothesis for arriving at a solution when the comparative term "less" was presented, as also supported by the mean scores and the below chance results. Unfortunately, this hypothesis was not the appropriate one. Children's hypothesis, for the "less"-type questions revolved on choosing the choice stimuli which was the equivalent of the smaller standard.

There was an observable difference in pattern of strategies for the Equalize and Compare questions with both comparative terms, as the interaction of question type by comparative term previously demonstrated. No Strategy was more evident in Compare questions than in Equalize questions regardless of comparative term. Furthermore, the Measuring Strategy was widely used in the "more" comparative term regardless of question type. Children seem to understand the question as a consequence of their understanding of the linguistic comparative term "more". In the same way, children seem to get confused as a consequence of their lack of understanding of the linguistic comparative term "less".

All previous findings in Experiment 2, on children's strategy use, were replicated in the present experiment.

4.4 Discussion

Performance on Compare word problems, in Experiment 3, is not affected by manipulations of display (that is, using the Undivided displays, or the "A1=B" displays, or the "A1≠B" displays). It was thought that the "A1=B" displays would considerably help children in solving problems of this nature, but the prediction was not supported by the results. These results lead one to believe that difficulties with part-whole relationships may have nothing to do with the Equalize and Compare difference. The hypothesis which was the basis for the present experiment was rejected by the negative results.

The Question by Comparative Term interaction was replicated from Experiment 2. However, this significant Equalize and Compare difference for the "more" comparative term, (but not for the "less"), did not appear to be true of the Undivided group, in this Experiment. It was expected that this would have then led to a Material x Question Type x Comparative Term interaction, which it did not. There are two hypotheses that may explain the consistent Question Type by Comparative Term interaction.

One hypothesis concerns spatial imagery. It seems that children are able to produce a spatial image with the Equalize-type question with the "more" comparative term where the lesser quantity can increase up to a certain amount. Children may find it more difficult to envisage a larger quantity decreasing. It is suggested that it is easier to imagine an increase in something than to cancel out something.

The second hypothesis concerns a linguistic effect. The evident floor effect with the "less" comparative term in the Undivided displays, may be due to the fact that "less" is a difficult word. (See Chapter Two.) It is speculated that the word "less" makes it too difficult for children to pursue the right strategies, even with Equalize problems.

Again, the same trends are observed regarding strategy choice as in Experiment 2. Children in the Undivided task, for the "more" comparative term, use the Measuring Strategy. However, this is only true for the Equalize-type questions, which they understand. This is only true, as well, for the Undivided displays. However, when presented with the "A1=B" and the "A1≠B" displays, children seem to prefer the Equivalence Strategy on the basis of height only. It seems they find the Equalize-type question too difficult in these other two displays; hence, utilizing the Equivalence Strategy, which they heavily rely on when they do not cognitively understand the question.

For the Compare-type question, as well as for both type questions in the "less" comparative term, for all three displays, children use an

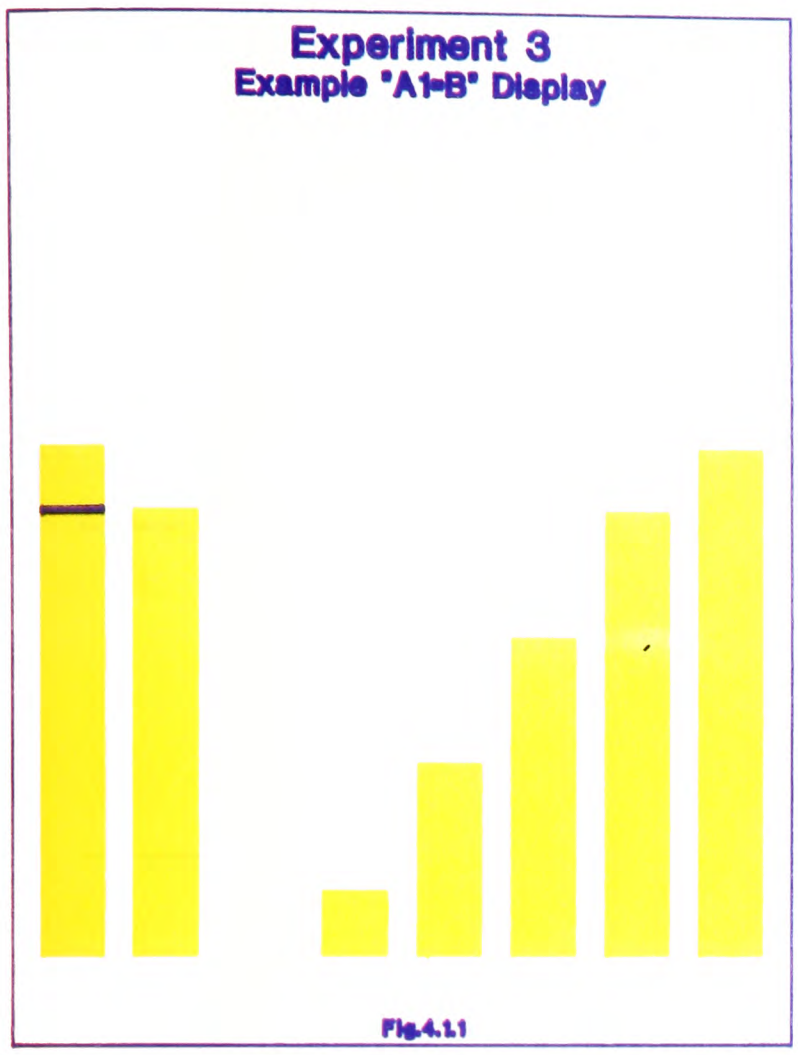
Equivalence Strategy whereby they choose the response on the basis of height. As was previously suggested, they seem to use this strategy when they find the word problem too difficult to handle.

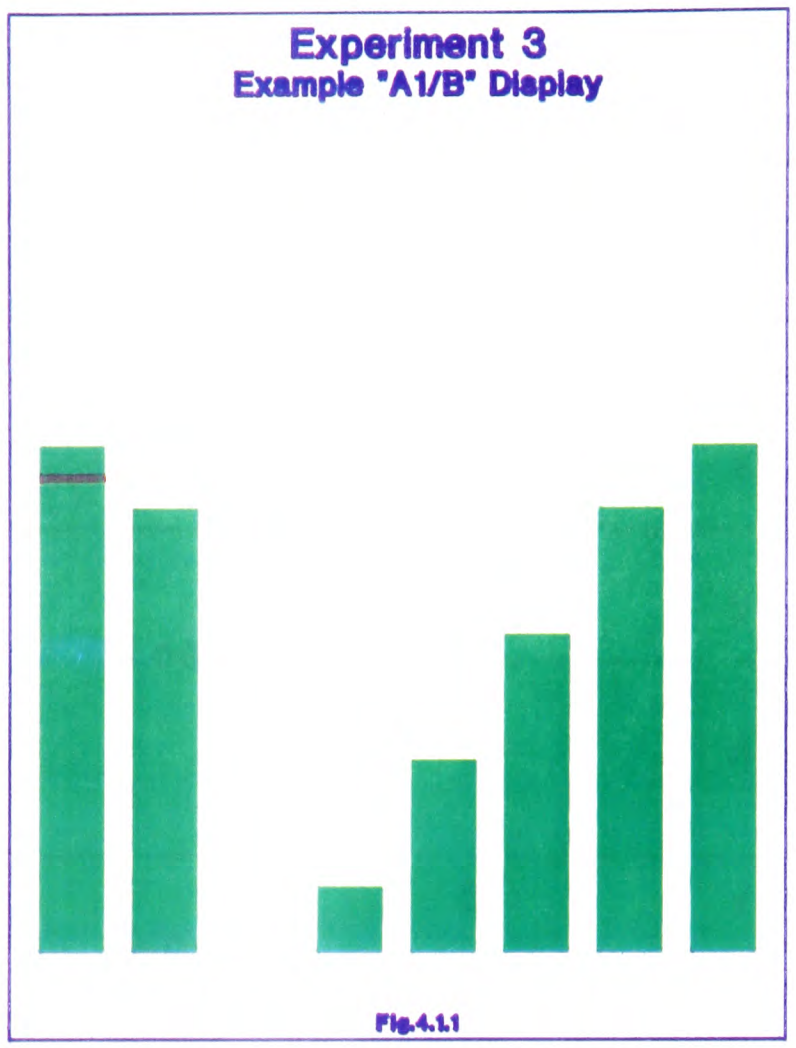
The same pattern of strategy use seems to be occurring throughout the three displays. Children exhibit the Equivalence Strategy in the "A1=B" display, which essentially establishes equivalence, as well as in the "A1≠B" condition, which essentially does not establish equivalence. This indicates that children, at this age of 5-to-7-years, still do not understand the concept of part-whole relationships. If they would have demonstrated some understanding they would have performed significantly better on the "A1=B" displays and significantly worse on the "A1≠B" displays. The Equivalence Strategy, as proposed in Experiment 2, indicates that the child is lacking an understanding about the parts that make up the whole. The results of this experiment clearly support this proposition.

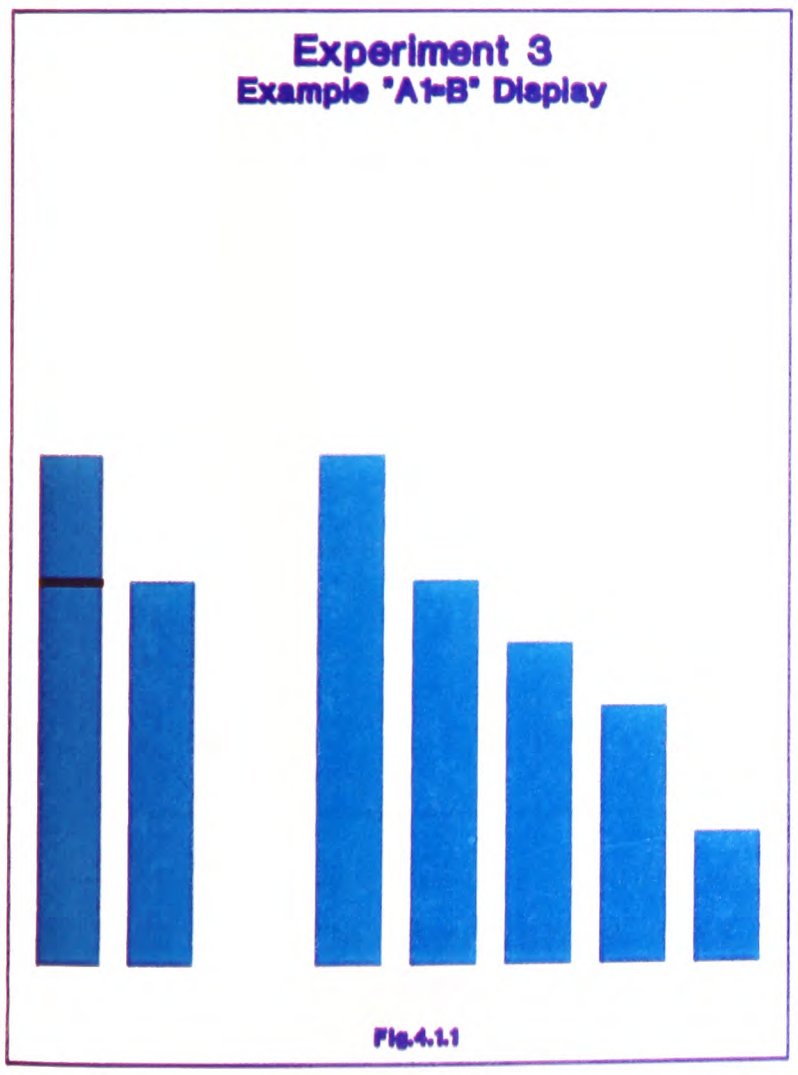
The fact that there were order effects showed that children had a tendency to get the "A1≠B" displays correct if the Compare question was presented first.

Experiment 4 was then designed to further look at the effects that different types of material could have on question type by investigating these in a different context. Experiment 4 was also designed to investigate whether the child's understanding of context is inseparable from the understanding of quantity itself.

FIGURES AND TABLES FOR EXPERIMENT 3


Experiment 3
Example "A1=B" Display
Fig.4.1.1


Experiment 3
Example "A1/B" Display
Fig.4.1.1


Experiment 3
Example "A1=B" Display
Fig.4.1.1

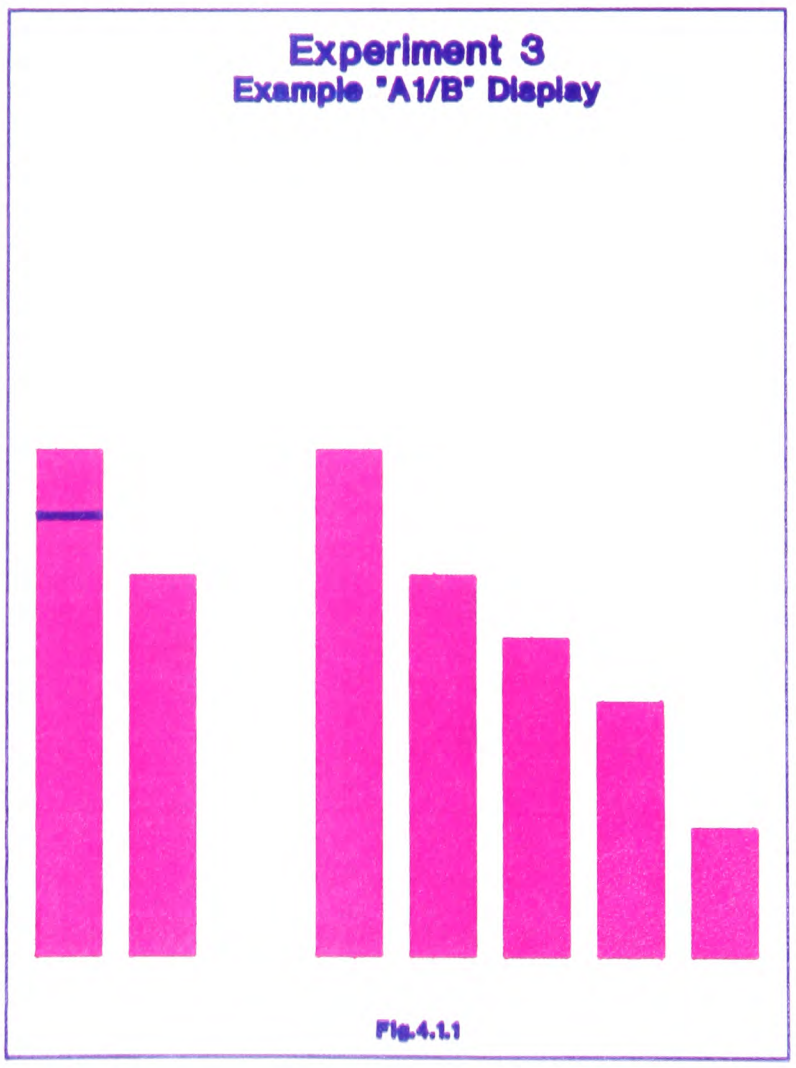

Experiment 3
Example "A1/B" Display
Fig.4.1.1

Table 4.2.1 The Design of Experiment 3

No. of children	Group	Comparative Term	No. of children	Question Type
16	UNDIVIDED	More	8	x
			8	y
16		Less	8	y
			8	x
16	"A1=B"	More	8	y
			8	x
16		Less	8	x
			8	y
16	"A1≠B"	More	8	x
			8	y
16		Less	8	y
			8	x
N=96				

x = Sequence Set Equalize first, Compare second
y = Sequence Set Compare first, Equalize second

Table 4.3.1 Means and Standard Deviations (out of 16) for the Number of Correct Responses

Material Group	Comparative Term	Question Type	
		Equalize	Compare
Undivided	More	7.31 (4.15)	7.25 (3.93)
	Less	1.75 (0.98)	1.38 (1.07)
"A1=B"	More	7.25 (4.60)	4.50 (4.06)
	Less	4.94 (5.17)	3.13 (2.82)
"A1≠B"	More	5.44 (4.80)	4.00 (5.18)
	Less	6.25 (3.87)	5.44 (3.75)

(Standard deviations in brackets.)

Table 4.3.2 Summary Table (ANOVA) for Total Correct Scores

Source	SS	DF	MS	F	PROB.
Mean	4880.333	1	4880.333	165.98	0.0000
Material[M]	5.635	2	2.818	0.10	0.9087
Ct	18.750	1	18.750	0.64	0.4268
Order[O]	7.521	1	7.521	0.26	0.6144
MCt	77.281	2	38.641	1.31	0.2742
MO	407.385	2	203.693	6.93	0.0016
CtO	31.688	1	31.688	1.08	0.3022
MCtO	7.531	2	3.766	0.13	0.8800
error	2469.875	84	29.403		
Question[Q]	143.521	1	143.521	19.09	0.0000
QM	10.760	2	5.380	0.72	0.4918
QCt	38.521	1	38.521	5.12	0.0262
QO	21.333	1	21.333	2.84	0.0958
QMCt	24.698	2	12.349	1.64	0.1996
QMO	10.635	2	5.318	0.71	0.4958
QCtO	36.750	1	36.750	4.89	0.0297
QMCtO	34.406	2	17.203	2.29	0.1077
error	631.375	84	7.516		

SS = Sum of Squared Deviations
DF = Degrees of Freedom
MS = Mean of Squared Deviations
F = Ratio of Variances
Prob. = Level of Significance

Experiment 3
QuestionxComparative Term

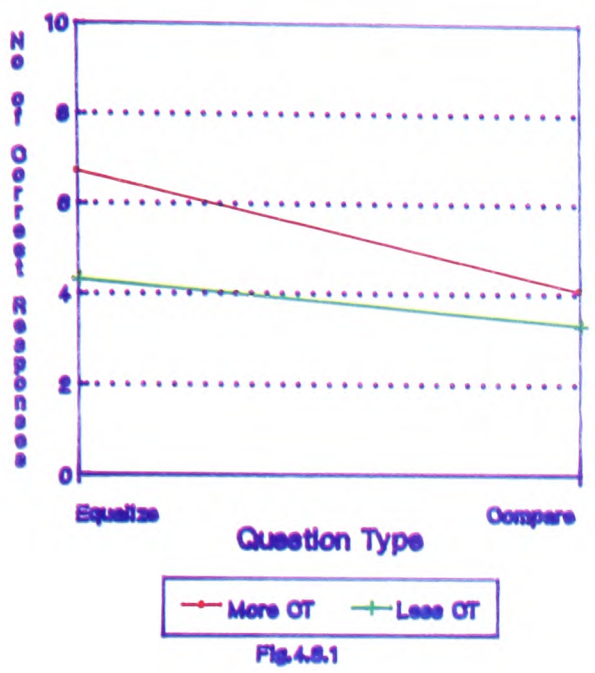


Fig.4.8.1

Experiment 3
Material x Order

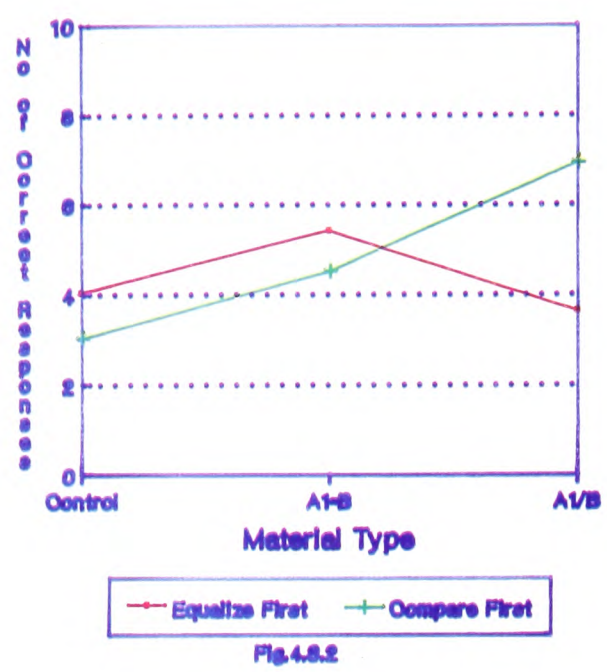


Fig.4.8.2

Experiment 3
QuestionxCtxOrder

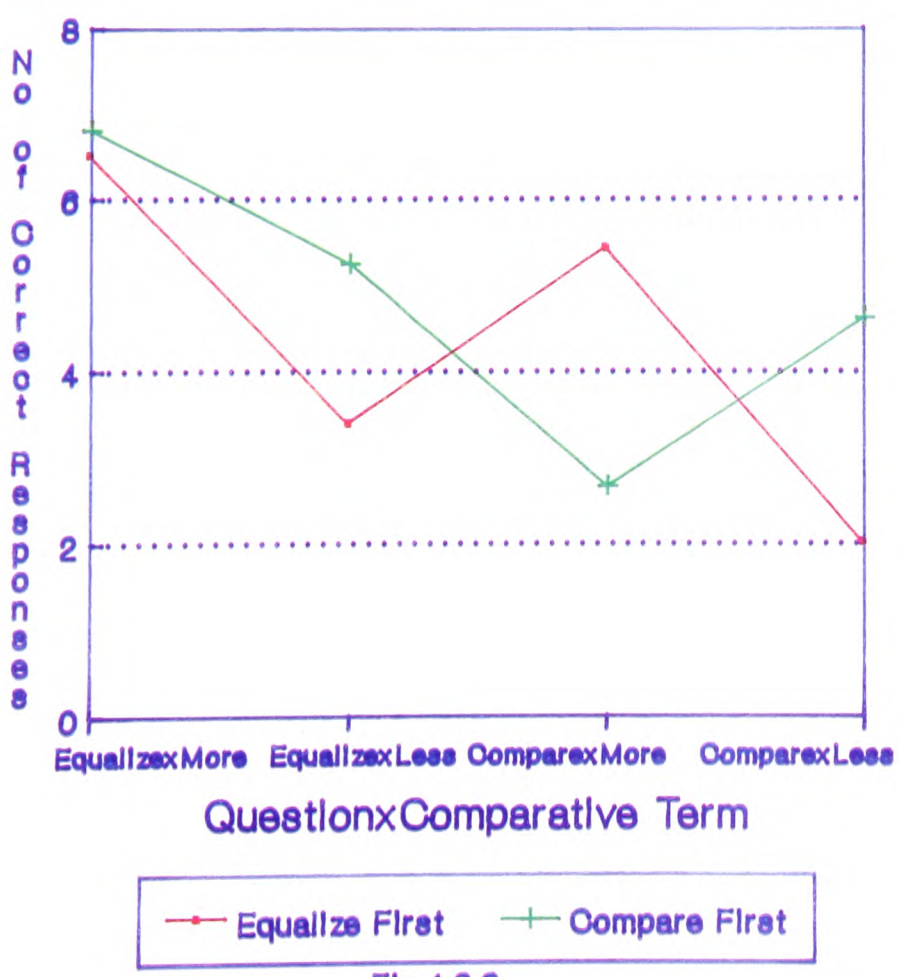


Fig.4.8.3

CHAPTER FIVE

STORY-TELLING AS CONTEXT FOR SOLVING MATHEMATICAL WORD PROBLEMS

5 Experiment 4 An Investigation of Responses made in Different Types of Question using Different Types of Material Depicting Quantity in Two Different Contexts

5.1 Introduction

Experiment 4 was designed to further look at the effects that different types of material could have on question type by investigating these in a different context. Hughes (1986) claimed that children's difficulties in learning mathematics stemmed from their lack of understanding of abstract concepts. As mathematics is abstract in its origin, children exhibit difficulty with this concept. According to Hughes, children's mathematical understanding will improve if it is put in some kind of context. Hence, Experiment 4 was designed to investigate whether the child's understanding of context is inseparable from the understanding of quantity itself by asking children Equalize and Compare questions in a story-telling context.

It was hypothesized that two possibilities could arise. Either,

(1) Children's overall performance would improve in both Equalize and Compare questions when presented in the story-telling context. This hypothesis would support Hughes' theory. Or,

(2) There would be a differential effect, where the story-telling context would help Equalize questions and children would continue to perform better in these than in Compare questions. As mentioned earlier (see Experiment 3), it is easier to imagine a change with Equalize problems, as they are

dynamic in nature, than it is with Compare problems, as they are static in nature. Putting Equalize problems in a story-telling context would contribute in making the change a more meaningful one, hence helping in the improvement of mentioned problems. However, putting Compare problems in a story-telling context, would not help in improving children's performance in mentioned problems, as the static essence of these remains the same.

Hence, Hughes would support hypothesis (1), where regardless of the nature of the two types of questions, children's performance would improve in both of them upon putting them in a story-telling context. This author would support hypothesis (2), where the nature of the two types of questions predisposes children's performance on them. This author's hypothesis will be further supported by observations on children's strategies, as previously examined in Experiments 2 and 3.

5.2 Method

(a) Subjects

One hundred and thirty-two children, ranging in age from 5;1 - 8;6 (mean age of 6;2), divided into eight groups (four groups of 17 children each and four groups of 16 children each) served as subjects. One group of 17 children ranging in age from 5;5 - 6;3 (mean age of 5;8), received the control discontinuous, comparative term "more" condition; another group of 17 children, ranging in age from 5;1 - 6;2 (mean age of 5;6), received the control discontinuous, comparative term "less" condition; another group of 17 children ranging in age from 5;3 - 6;3 (mean age of 5;9), received the control continuous, comparative term "more" condition; and the final group of 17 children ranging in age from 5;1 - 6;3 (mean age of 5;6), received the control continuous, comparative term "less" condition. One group consisted of 16

children ranging in age from 5;3 - 8;6 (mean age of 6;9), received the story discontinuous, comparative term "richer" condition; another group of 16 children, ranging in age from 5;2 - 7;6 (mean age of 6;3), received the story discontinuous, comparative term "poorer" condition; another group of 16 children ranging in age from 5;4 - 7;8 (mean age of 6;7), received the story continuous, comparative term "taller" condition; and the final group of 16 children ranging in age from 6;1 - 7;8 (mean age of 6;9), received the story continuous, comparative term "shorter" condition. The children were all attending a first school in Oxford, England.

The four groups of 17 children each, receiving the control conditions, were the same subjects as in Experiment 2. In fact, Experiment 2 was compared against the story-telling conditions provided in this Experiment 4.

Seventy-three additional children, ranging in age from 5;2 - 8;8 (mean age of 6;6) divided into four groups (two groups of 24 children each, one group of 20 children, and one group of five children) served as pilot subjects. One group consisted of 24 children, ranging in age from 7;3 - 8;8 (mean age of 8;1); another group consisted of 24 children, ranging in age from 5;2 - 6;8 (mean age of 5;8); another group consisted of 20 children, ranging in age from 5;2 - 6;5 (mean age of 5;8); and the final group consisted of 5 children, ranging in age from 6;2 - 7;8 (mean age of 6;7). These pilot subjects were given two standardized tests: the short form of the British Picture Vocabulary Scale (BPVS) and the arithmetic section of the Weschler Intelligence Scale for Children Revised (WISC-R), as well as a test devised for this experiment, including four addition problems and four subtraction problems (see Appendix 3). The children were all attending a first school in Oxford, England.

(b) Apparatus

The same apparatus was used as for Experiments 2 and 3.

(c) **Design**

Each child was given two sequences, in one session, in which there were 16 trials on each sequence. One sequence of 16 trials was for the Equalize-type questions and the other sequence of 16 trials was for the Compare-type questions. Each of the 16 trials had a different presentation. There were 16 sequences of the 16 displays for the Story condition. Each child received one sequence for each question-type and the 17th child got a random sequence. Each different display had a trial label, represented by a number. Each sequence consisted of a randomized order of trials. These were the following:

1st sequence:	15-9-5-1-11-7-4-12-8-6-3-2-14-13-10-16
2nd sequence:	1-13-9-12-6-3-11-14-16-10-15-7-2-5-4-8
3rd sequence:	2-10-15-16-12-11-6-14-5-7-1-4-9-13-8-3
4th sequence:	10-16-8-13-12-5-4-1-15-14-9-2-6-11-3-7
5th sequence:	10-6-14-16-2-3-15-9-5-7-11-4-1-12-8-13
6th sequence:	10-8-15-3-9-6-14-12-1-11-2-7-16-5-4-13
7th sequence:	13-16-15-14-7-12-6-2-11-4-8-5-10-3-1-9
8th sequence:	4-7-8-16-15-5-10-9-6-3-1-12-13-14-11-2
9th sequence:	14-11-9-2-3-10-6-1-4-12-16-5-13-7-8-15
10th sequence:	14-3-4-16-1-7-13-5-15-6-2-12-10-9-11-8
11th sequence:	12-6-13-16-10-2-7-5-8-4-9-11-14-1-3-15
12th sequence:	8-15-9-16-2-7-4-12-6-10-3-5-13-14-11-1
13th sequence:	16-4-8-5-10-3-15-11-7-14-1-6-13-9-12-2
14th sequence:	10-5-4-3-9-8-11-7-16-1-2-12-6-15-14-13
15th sequence:	2-7-4-16-10-9-12-11-1-8-15-3-5-13-6-14
16th sequence:	9-3-15-10-6-8-7-12-1-4-11-14-13-16-5-2

There were also 16 different sequences of the 16 displays for the Control condition. These were the following:

1st sequence: 15-2-5-12-7-3-6-10-8-14-1-13-9-11-4-16
2nd sequence: 4-14-10-11-15-2-1-12-6-5-9-3-13-16-8-7
3rd sequence: 14-4-8-7-10-6-5-16-2-13-15-11-9-1-12-3
4th sequence: 15-14-1-10-8-4-13-11-12-6-3-7-5-9-2-16
5th sequence: 5-1-6-12-9-2-14-16-13-15-7-8-3-4-10-11
6th sequence: 1-13-5-14-11-10-4-15-12-3-9-16-6-7-8-2
7th sequence: 4-12-16-6-7-13-11-14-8-1-5-2-3-15-9-10
8th sequence: 15-12-13-4-3-10-14-8-9-7-6-1-16-11-5-2
9th sequence: 11-1-14-8-12-10-7-13-15-6-4-9-2-16-5-3
10th sequence: 13-14-9-4-6-11-12-3-2-8-7-15-1-10-16-5
11th sequence: 3-15-8-4-11-7-14-6-9-5-13-16-12-1-10-2
12th sequence: 2-6-8-10-13-11-5-16-7-9-14-3-1-12-15-4
13th sequence: 13-9-11-2-5-1-10-7-6-4-15-3-8-12-14-16
14th sequence: 5-11-4-13-7-6-10-16-8-1-9-15-2-12-14-3
15th sequence: 13-14-5-16-6-7-11-12-10-8-4-15-9-3-1-2
16th sequence: 11-13-8-5-4-14-3-7-9-6-15-12-10-16-2-1

The question types were presented in alternating order. Half of the children in one condition received the Equalize-type question first followed by the Compare-type question (sequence-set X); the other half were given the task in the reverse order (sequence-set Y).

Hence, a mixed design with repeated measures on Conditions (5: Story [S], Material [M], Comparative Term [CT], Order [O], Question Type [Q]) was used. Groups (Eight groups: four of 17 children each: (1) Control discontinuous material, comparative term "more", (2) Control discontinuous material, comparative term "less", (3) Control continuous material, comparative term "more", (4) Control continuous material, comparative term "less"; and four of 16 children each: (5) Story discontinuous material, comparative term "richer", (6) Story discontinuous material, comparative term "poorer", (7) Story continuous material, comparative term "taller", (8) Story

continuous material, comparative term "shorter") was the Between factor. The design for Experiment 4 is illustrated in Table 5.2.1.

(d) Materials

The same computer-driven displays were used as for Experiment 2 (see Section 3.2(d)).

The story-telling condition, however, was a more elaborate one. It consisted of the same two questions (Equalize and Compare) with the comparative terms, "richer" or "poorer", for the discontinuous material, and the comparative terms, "taller" or "shorter", for the continuous material. It also consisted of a preamble to the questions. This preamble consisted of a short story before the questions were asked. The story was composed of two or three sentences of relevant information. The story emphasized the saliency of the characteristics, "taller" or "shorter", and "richer" or "poorer". In the discontinuous material, for example, the boy was described as having lost money and being so much poorer than the girl; the girl was described as having gained money and being so much richer than the boy. In the continuous material, for example, the boy was described as having shrunk and being so much shorter than the girl; the girl was described as having grown and being so much taller than the boy.

(e) Procedure

Children were taken randomly from their classes individually. They were shown each display one at a time as displayed on the screen. Each of the 16 illustrations was displayed twice under the condition: (1) Question Type (Equalize or Compare).

For each display, a child was asked two questions (16 questions in the Equalize form and 16 questions in the Compare form). Hence, each child was asked a total of 32 questions in all.

In all the displays, the larger standard of the comparative pair was on the far left. For each of the 16 displays in the control discontinuous condition, the experimenter asked one Compare-type question of the form, on one occasion, *"How many more does this one* (pointing to the larger standard) *have than this one* (pointing to the smaller standard)*?"*; and one Equalize-type question of the form, *"How many more does this one* (pointing to the smaller standard) *need to have, to have the same as this one* (pointing to the larger standard)*?"*, on another occasion. For each display in the control continuous condition, the experimenter asked one Compare-type question of the form, *"How much more does this one* (pointing to the larger standard) *have than this one* (pointing to the smaller standard)*?"*; and one Equalize-type question of the form, *"How much more does this one* (pointing to the smaller standard) *need to have, to have the same as this one* (pointing to the larger standard)*?"*. The child then selected an answer (one of the five stimuli from the choice stimuli).

In the other comparative group, a child was asked again two questions for each display in the control discontinuous condition: one of the Compare-type form, *"How many less does this one* (pointing to the smaller standard) *have than this one* (pointing to the larger standard)*?"*; and one of the Equalize-type form, *"How many less does this one* (pointing to the larger standard) *need to have, to have the same as this one* (pointing to the smaller standard)*?"*. For each display in the control continuous condition, the experimenter would ask one Compare-type question of the form, *"How much less does this one* (pointing to the smaller standard) *have than this one* (pointing to the larger standard)*?"*; and one Equalize-type question of the form, *"How much less does this one* (pointing to the larger standard) *need to*

have, to have the same as this one (pointing to the smaller standard)*?"*. The child then selected an answer (one of the five stimuli from the choice stimuli).

For the story condition (i.e. the story-telling condition), the term "richer" was substituted for "more" in the story discontinuous condition. The child was asked two questions for each display: one of the Compare-type form, *"How much richer is this one* (pointing to the larger standard) *than this one* (pointing to the smaller standard)*?"*; and one of the Equalize-type form, *"How much richer does this one* (pointing to the smaller standard) *have to get, to have the same as this one* (pointing to the larger standard)*?"*.

The other version of this task had the term "poorer" substituted for "less". The child was asked two questions for each display in the story discontinuous condition: one of the Compare-type form, *"How much poorer is this one* (pointing to the smaller standard) *than this one* (pointing to the larger standard)*?"*; and one of the Equalize-type form, *"How much poorer does this one* (pointing to the larger standard) *have to get, to have the same as this one* (pointing to the smaller standard)*?"*.

In another task, the term "taller" was substituted for "more". The child was asked two questions for each display in the story continuous condition: one of the Compare-type form, *"How much taller is this one* (pointing to the larger standard) *than this one* (pointing to the smaller standard)*?"*; and one of the Equalize-type form, *"How much taller has this one* (pointing to the smaller standard) *to grow, to be the same as this one* (pointing to the larger standard)*?"*.

The other version of this task had the term "shorter" substituted for "less". The child was asked two questions for each display in the story continuous condition: one of the Compare-type form, *"How much shorter is this one* (pointing to the smaller standard) *than this one* (pointing to the larger standard)*?"*; and one of the Equalize-type form, *"How much shorter has this*

one (pointing to the larger standard) *to shrink, to be the same as this one* (pointing to the smaller standard)*?*".

All the responses were recorded and stored in the computer.

5.3 Results

The following section will be divided into two sub-sections. One sub-section will be based on the pattern of correct and incorrect responses; the other sub-section will be based on the different observed strategies that children used to arrive at their responses.

5.3.1 Comparing Correct Scores across Conditions

The principal data were the number of Equalize questions answered correctly, and the number of Compare questions answered correctly. The mean number of correct responses for each type of question is shown in Table 5.3.1. These means were obtained from scores on the 16 questions on each question type and give an overall picture of children's performance. In general, once again, more correct responses were given to the Equalize questions than to the Compare questions.

Table 5.3.1 Means and Standard Deviations (out of 16) for the Number of Correct Responses

The means and standard deviations for the total number of correct responses made in each condition by the children were given in Table 5.3.1. This table shows that Equalize questions were superior to Compare questions with discontinuous material and with continuous material when this was presented with the comparative term "more". In particular, Equalize questions were twice as good as Compare questions when given with

discontinuous material in the "more" comparative term. Overall, there was a difference between Equalize and Compare questions, as most correct responses were made to the former and not the latter. However, in one of these cells, there was no difference in correct responses between Equalize- and Compare-type questions, as the children obtained such a low score. As in Experiment 2, this occured when these were presented with control continuous material in the "less" comparative term. The results replicated those of Experiment 2.

Calculating the Probability of Correct Responses occurring purely by Chance

As for Experiment 2 and 3, children had a choice of five responses (from the choice stimuli). Therefore, there was a 20% chance of their choosing any one of these responses. Children were considered to be performing at chance level if they got a total of 3.2 correct responses. The probability that the correct responses occured simply by chance was explored using a modification of the standard hypothesis testing procedure with the means in Table 5.3.1. (See Section 3.3.1.)

Scores significantly above chance

The proportions of correct responses made in all, but four of the cell means, were significantly ($p<0.01$) greater than expected purely by chance. The proportions of correct responses made by the group to the Compare question with control continuous material, comparative term "more" was significantly ($p<0.05$) greater than expected purely by chance. The proportions of correct responses made by the group to the Compare question with story discontinuous material, comparative term "poorer" was significantly ($p=0.05$) greater than expected purely by chance. This indicates that the children were not acting randomly.

Scores at chance

The proportion of correct responses made by two groups: (1) Equalize question with control continuous material, comparative term "less", and (2) Compare question with story discontinuous material, comparative term "richer", were not significantly different from chance. This indicates that for these groups, children were acting randomly.

5.3.2 Analysis of Variance

These results were analyzed by subjecting the raw scores for the total number of correct responses to a 2x2x2x2x2 ANOVA. The main terms were: (1) Story (control or story), (2) Material (discontinuous or continuous), (3) Comparative Term (more or less), (4) Order (Equalize questions first, Compare second or Compare question first, Equalize second), and (5) Question Type (Equalize or Compare) with repeated measures on the last factor. The results of this analysis are presented in Table 5.3.2.

The significant main effect of Question ($F(1,116)=45.46$, $p<0.001$) indicated that Equalize questions once again were easier than Compare questions.

The significant interaction between Question Type and Comparative Term ($F(1,116)=5.30$, $p<0.05$), showed that the effects of the Comparative Term affect significantly the children's performance on Question Type. This interaction was explored using a Newman-Keuls Multiple Range Test. It was found that the comparative terms "more", "richer", and "taller", produce a significantly higher number of correct responses from the children on the Equalize-type questions than on the Compare-type questions. Furthermore, Equalize questions with these more comparative terms are significantly different from Compare questions with the "less", "poorer", and "shorter" comparative terms. Children are just as bad on Equalize-type questions as on Compare-type questions when these are presented with the less

comparative terms. Hence, the Equalize and Compare difference is more evident with the more comparative terms than with the less comparative terms. This result replicated those of Experiments 2 and 3. A graphical representation ot this comparison is presented in Figure 5.3.1.

The significant interaction between Question Type and Material ($F(1,116)=5.41$, $p<0.05$), showed that the effects of Material affect significantly the children's performance on Question Type. This interaction was explored using a Newman-Keuls Multiple Range Test. It was found that children perform significantly better on Equalize questions than on Compare questions when discontinuous material is used. Children are helped significantly by the use of discontinuous material in Equalize problems, but not in Compare problems. This result replicated that one of Experiment 2. A graphical representation of this comparison is presented in Figure 5.3.2.

The significant interaction between Question Type and Order ($F(1,116)=4.04$, $p=0.05$), showed that the effect of order presentation affects significantly the children's performance on Question Type. This interaction was explored using a Newman-Keuls Multiple Range Test. Children perform significantly better on the Equalize questions regardless of the order of presentation than on the Compare questions when these are asked first followed by the Equalize questions. It so seems that children find the most difficult-type question even more difficult when it precedes the Equalize-type question. A graphical representation of this comparison is presented in Figure 5.3.3.

The significant interaction between Material and Story ($F(1,116)=6.44$, $p<0.05$), showed that the effect of Material affects significantly the children's performance on Story. A Newman-Keuls Multiple Range Test applied to this interaction demonstrated that each condition did not differ significantly from each other. A graphical representation of this comparison is presented in Figure 5.3.4.

5.3.3 Observing Children's Strategies

Observations were again recorded on children's strategies in making their quantity judgements. One more strategy was added to the same four categories of responses recorded in Experiments 2 (see Section 3.3.2) and 3 (see Section 4.3.2).

The Extreme Choice Stimuli Strategy entailed choosing the largest item among the choice stimuli when asked either type of question with the more comparative terms, or choosing the smallest item among the choice stimuli when asked either type of question with the less comparative terms. This new classification surged from observing children's errors. As observed previously in Experiment 2 (see Section 3.3.3), children were making this error consistently, thinking it was the correct response and thinking they had a valid explanation for it. Though systematically wrong, they had a definite hypothesis when performing the task. Hence, the Extreme Choice Stimuli Error becomes more of a strategy used to arrive at the correct solution.

Percentage Number of Responses for each Strategy Type was recorded for both Equalize and Compare questions.

Strategy Types	Equalize Questions	Compare Questions
Counting Strategy	31%	15%
Measuring Strategy	27%	15%
Equivalence Strategy	33%	48%
Extreme Strategy	3%	12%
No Apparent Strategy	6%	10%

It was found, as in Experiments 2 and 3, that for both Equalize and Compare questions, children prefered to use the Height Equivalence Strategy the great majority of the time. However, there was also a different pattern for the Equalize questions. The Counting Strategy was more evident in Equalize

questions than in Compare. It is also worth noting that the Measuring Strategy was widely used in the Equalize-type question and not so much so in the Compare-type question. The Measuring Strategy is a substitute for the Counting Strategy, for continuous material. Children seem to use it more frequently, as they use the Counting Strategy for discontinuous material, when they understand the question.

Percentage Number of Responses for each Strategy Type was recorded for both discontinuous and continuous material.

Strategy Types	Discontinuous Material	Continuous Material
Counting Strategy	37%	10%
Measuring Strategy	2%	39%
Equivalence Strategy	39%	42%
Equivalence/Counting	17%	6%
Equivalence/Height	22%	36%
Extreme Strategy	10%	5%
No Apparent Strategy	12%	4%

It was found, as in Experiment 2 and 3, that for both discontinuous and continuous material, children prefered to use the Height Equivalence Strategy, as found previously for the question types. There was a further finding in this type of condition. The Counting Strategy was widely used for the discontinuous material and the Measuring Strategy was widely used for the continuous material.

Percentage Number of Responses for each Strategy Type was recorded for both comparative terms: "more" and "less".

Strategy Types	More Comparative Term	Less Comparative Term
Counting Strategy	27%	19%
Measuring Strategy	24%	18%
Equivalence Strategy	35%	46%
Equivalence/Counting	10%	13%
Equivalence/Height	25%	33%
Extreme Strategy	9%	6%
No Apparent Strategy	5%	11%

It was found, as in Experiment 2 and 3, that for both comparative terms, children prefered to use the Height Equivalence Strategy, as found previously for the question types and for display types. Children tended to choose the equivalent of the larger standard for the "more"-type questions, and/or the equivalent of the smaller standard for the "less"-type questions to arrive at a solution. Both the Counting and the Measuring Strategy were widely used for the comparative term "more". This was also true for the Equalize-type questions. It is interesting that these two variables triggered widespread use of the Counting and the Measuring Strategy. Furthermore, it is interesting to note that there was an interaction of Question Type and Comparative Term. It is speculated that the this interaction triggered use of this strategy type.

Percentage Number of Responses for each Strategy Type was recorded for both Control and Story conditions (i.e. non-story telling or story telling).

Strategy Types	Control Condition	Story Condition
Counting Strategy	22%	23.5%
Measuring Strategy	15%	22.5%
Equivalence Strategy	54%	37%
Extreme Strategy	N/A	9%
No Apparent Strategy	9%	8%

It was found that for both the Control and Story Condition, children prefered to use the Height Equivalence Strategy the great majority of the time. However, this was more evident for the Control Condition than for the Story-telling Condition.

All previous findings in Experiment 2 and 3, on children's strategy use, were replicated in the present experiment.

5.4 Discussion

This experiment was set out to answer two questions.:

(1) Would there be an Equalize and Compare difference, in general?

(2) Would this Equalize and Compare difference be different in the story-telling context than in the Control context?

Answers to these questions were sought in an attempt to explain whether Hughes' (1986) hypothesis concerning the possibility of children's overall performance in both Equalize and Compare questions improving when presented in the story-telling context was right, or whether this author's hypothesis concerning a differential effect where the story-telling context

would help children's performance in Equalize problems more than in Compare problems was right. A highly significant main effect for story would have resulted should Hughes' hypothesis be right. A significant cross-over interaction between story and question would have resulted should this author's hypothesis be right.

The results of Experiment 4 indicated that children perform well on the type of problem which requires them to compare quantities, as they find it difficult to solve Compare-type word problems, but find it easier to solve Equalize-type word problems reasonably well, particularly if these are presented in discontinuous fashion and using the "more" comparative term. These results replicated those of Experiment 2.

As in Experiment 2, Equalize problems were difficult with "less". This author hypothesizes that it is the nature of the problem that explains this result. Children have demonstrated to be better on dynamic problems when these involve a change-up, rather than when these involve a change-down, as it is easier to imagine an increase and harder to imagine a decrease.

Performance on Compare word problems was not affected in Experiment 4, as in Experiment 2, by manipulations either of comparative terms (that is, using "more" or "less") or of material (that is, using discontinuous or continuous material). However, using different type of material (discontinuous), affected children's performance on the Equalize word problems, as well as, using different comparative markers ("more", "richer", or "taller") affected children's performance on Equalize word problems.

Children also prefered the story condition only when continuous material was used, and prefered the non-story condition when discontinuous material was used. It seems that stories aid children with material they find difficult, such as continuous material. However, stories do not aid children

with material with which they are relatively successful, such as discontinuous material.

The lack of a significant cross-over interaction between story and question did not confirm this author's hypothesis. Moreover, Hughes' hypothesis was not supported by the results. The negative results have demonstrated that a story-telling context does not aid Equalize questions more than Compare questions.

It seems that children in the continuous task, for the "more" comparative term, use an equivalent strategy to that of counting in the discontinuous task for the "more" comparative term. This equivalent strategy is a measuring one. However, as previously recorded, this is only true for the Equalize-type question, which they understand.

However, for the Compare-type question, as well as for both type question in the "less" comparative term, children use an Equivalence Strategy whereby they choose the response on the basis of counting for discontinuous material and on the basis of height for continuous material. They seem to use this strategy when they find the word problem too difficult to handle.

However, when questions are posed in a story-telling context, children seem to use different strategies to solve the problems. For the discontinuous material with both the "richer" and "poorer" comparative terms, children use the Counting Strategy most effectively, but this is only evident in the Equalize-type questions. For the Compare-type questions, children prefer to use the Equivalence by Height Strategy. They find this Compare-type question difficult to handle for both comparative terms. They do not even demonstrate use of the Equivalence Strategy by Counting, regardless of the material being presented in a discontinuous fashion, hence cueing them to count. Strategies in this story-telling context are triggered only by the type of question presented and not by the comparative terminology.

For the continuous material, posed in a story-telling context, children prefer to use the Measuring Strategy for Equalize-type questions with both comparative terms, "taller" and "shorter", and for the Compare-type question with the comparative term "taller". For the Compare-type question with the comparative term "shorter", children prefer to use the Equivalence by Height strategy. Children demonstrate a greater understanding of Equalize and Compare questions when posed in a story-telling context with continuous material, as demonstrated by the strategies used.

The strategies used are consistent with the results obtained. The success in the usage of pertinent strategies is reflected by the children's attempt to count in the Equalize-type question when discontinuous material is presented with the "poorer" comparative term, and by the children's attempt to measure in the Equalize-type question when continuous material is presented with the "shorter" comparative term, as well as for the Compare-type question with the "taller" comparative term.

The cognitive model developed in Experiment 2 (see Figure 3.4.1), is also applicable in a story-telling context.

However, this sequence is not only noticeable for Equalize-type questions, as was previously found in Experiments 2 and 3. For Compare-type questions, children no longer seem to stumble on the Equivalence Strategy of this sequence, which also previously triggered exactly the same reaction when the comparative term, "less", was used. When the questions are given in a story-telling context, children are more successful at arriving at the most refined strategy for the Equalize-type questions with the less comparative markers and for the Compare-type question, too, with the more comparative markers, when presented with continuous material.

The Equivalence Strategy, known as the transition ground the child must travel from a No Apparent Strategy to a Measuring Strategy or a Counting Strategy, is less used in the story-telling context. The fact that the

child demonstrates less use of this strategy indicates that the child has a better grasp of the concept of absolute size, as the child begins to measure or count more frequently.

It should also be noted that counting would not always yield the correct response. Some children would attempt to use a counting-all strategy when solving these addition and subtraction problems. Obviously, this strategy was not the proper one to use for these Equalize- and Compare-type problems. Some children who appropriately used a COS strategy (counting-on from the smaller addend), which would, in principle, yield the correct answer, would still not achieve it. The children would begin by counting the cardinal value of the smaller addend and continue on from there (i.e. 4-2=X, "2,3,4 = 3!).

The appearance of the Extreme Choice Stimuli Strategy, in the Compare-type questions in the story-telling context, may demonstrate that children do not fully comprehend this type of question. For some reason, some children interpret the Compare-type question as asking them to determine which is the richest/tallest and/or which is the poorest/shortest. It is noteworthy that this strategy was only evident in the story-telling context and predominantly evident for the Compare-type questions. Piaget and Szeminska (1952) had also found that when children were asked, "Which has more?", they would tend to interpret the question as, "Which one has the most?". These results provide only a little support for the results of Piaget. The percentage of children who demonstrated use of this strategy was not a great one, 10% for discontinuous material and 5% for continuous material.

Furthermore, the success at Compare-type questions was attributed to be dependent on whether or not they were presented with the Equalize-type questions first. This was never the case vice-versa. It so seems that this effect can go hand in hand with Gelman's (1982) training studies, as well as a further step forward from Starkey and Cooper's (1980) habituation experiments. It seems that children, when accustomed to Equalize-type

questions, habituate to them and are then able to perform successfully on Compare-type questions. The effect can be parallel to one of training. This has strong educational implications as training children on Equalize-type questions can then trigger successful performance on Compare-type questions.

FIGURES AND TABLES FOR EXPERIMENT 4

Table 5.2.1 The Design of Experiment 4

No. of children	Group	Comparative Term	No. of children	Question Type (Seq.Set)
17	Control Discontinuous	More	8 --- 9	x --- y
17		Less	9 --- 8	y --- x
17	Control Continuous	More	9 --- 8	y --- x
17		Less	8 --- 9	x --- y
16	Story Discontinuous	Richer	8 --- 8	x --- y
16		Poorer	8 --- 8	y --- x
16	Story Continuous	Taller	8 --- 8	y --- x
16		Shorter	8 --- 8	x --- y

N = 132

x = Sequence Set Equalize first, Compare second
y = Sequence Set Compare first, Equalize second

Table 5.3.1 Means and Standard Deviations (out of 16) for the Number of Correct Responses

Group	Comparative Term	Question Type: Equalize	Question Type: Compare
Control Discontinuous	More	10.48 (4.36)	5.08 (5.74)
	Less	7.96 (4.73)	5.25 (4.48)
Control Continuous	More	7.91 (4.71)	4.43 (5.39)
	Less	4.27 (2.94)	4.84 (3.91)
Story Discontinuous	Richer	7.00 (5.77)	2.81 (3.78)
	Poorer	7.75 (5.83)	4.31 (4.66)
Story Continuous	Taller	9.50 (5.04)	6.88 (6.03)
	Shorter	7.63 (4.27)	5.50 (4.32)

(Standard deviations in brackets.)

Table 5.3.2 Summary Table (ANOVA) for Total Correct Scores

Source	SS	DF	MS	F	PROB.
Mean	10612.556	1	10612.556	297.95	0.0000
Material[M]	0.101	1	0.101	0.00	0.9577
Story[S]	1.417	1	1.417	0.04	0.8423
ComparativeTerm[Ct]	44.485	1	44.485	1.25	0.2661
Order [O]	72.744	1	72.744	2.04	0.1557
MS	229.504	1	229.504	6.44	0.0125
MCt	41.711	1	41.711	1.17	0.2814
SCt	21.542	1	21.542	0.60	0.4383
MO	4.985	1	4.985	0.14	0.7090
SO	7.276	1	7.276	0.20	0.6521
CO	27.383	1	27.383	0.77	0.3824
MSCt	22.068	1	22.068	0.62	0.4328
MSO	82.417	1	82.417	2.31	0.1309
MCtO	3.697	1	3.697	0.10	0.7479
SCtO	11.476	1	11.476	0.32	0.5714
MSCtO	21.804	1	21.804	0.61	0.4356
error	4131.813	116	35.619		
Question[Q]	563.673	1	563.673	45.46	0.0000
QM	67.092	1	67.092	5.41	0.0217
QS	1.848	1	1.848	0.15	0.7002
QCt	65.714	1	65.714	5.30	0.0231
QO	50.101	1	50.101	4.04	0.0467
QMS	5.567	1	5.567	0.45	0.5042
QMCt	1.286	1	1.286	0.10	0.7480
QSCt	31.036	1	31.036	2.50	0.1163
QMO	1.351	1	1.351	0.11	0.7420
QSO	23.411	1	23.411	1.89	0.1721
QCtO	0.486	1	0.486	0.04	0.8434
QMSCt	2.693	1	2.693	0.22	0.6421
QMSO	11.861	1	11.861	0.96	0.3301
QMCtO	4.985	1	4.985	0.40	0.5273
QSCtO	0.679	1	0.679	0.05	0.8154
QMSCtO	5.435	1	5.435	0.44	0.5092
error	1438.257	116	12.399		

SS = Sum of Squared Deviations
DF = Degrees of Freedom
MS = Mean of Squared Deviations
F = Ratio of Variances
Prob. = Level of Significance

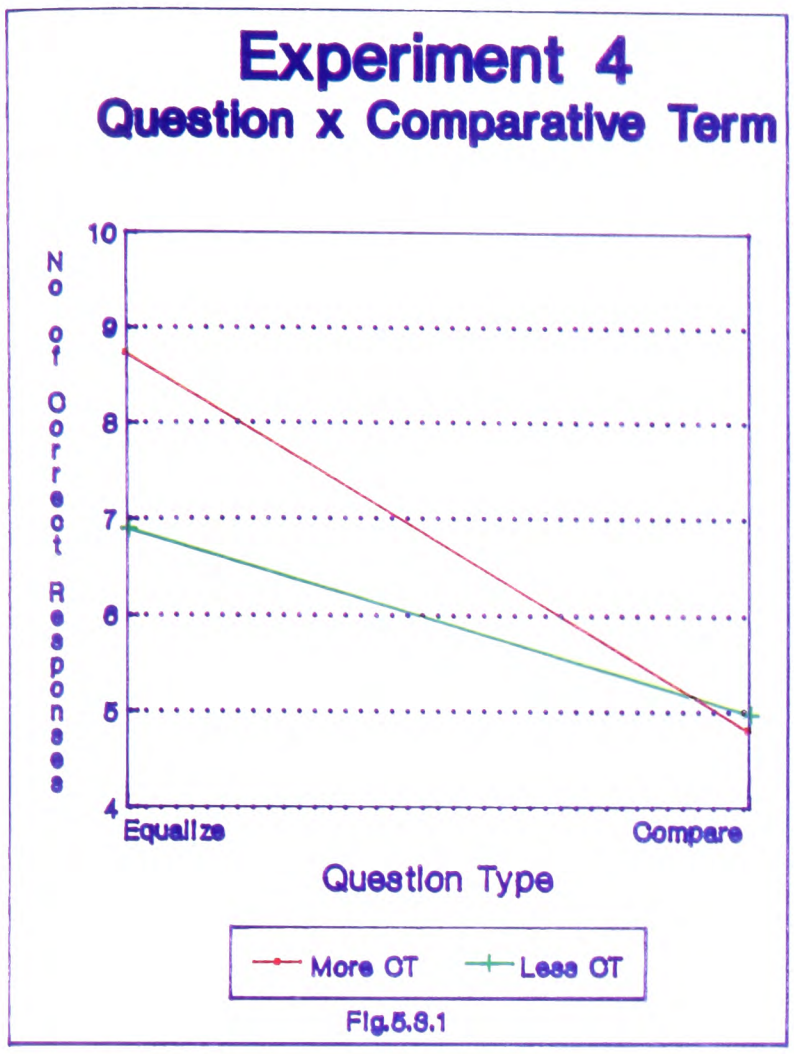


Fig.5.8.1

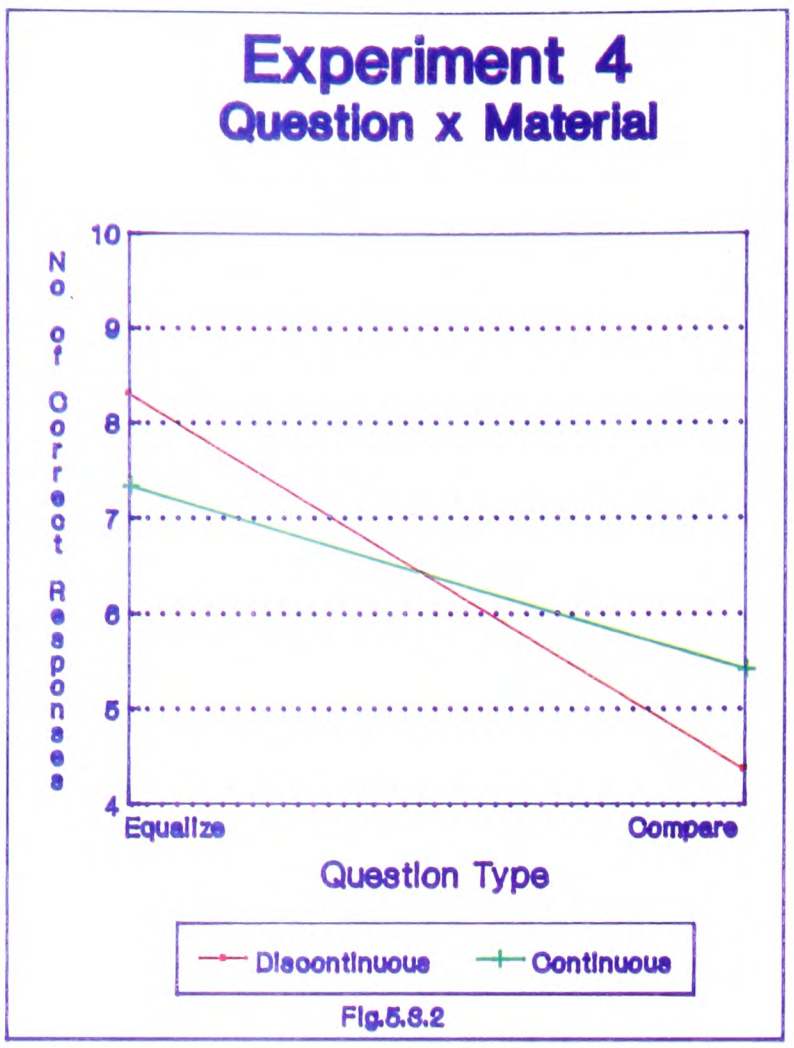


Fig.5.8.2

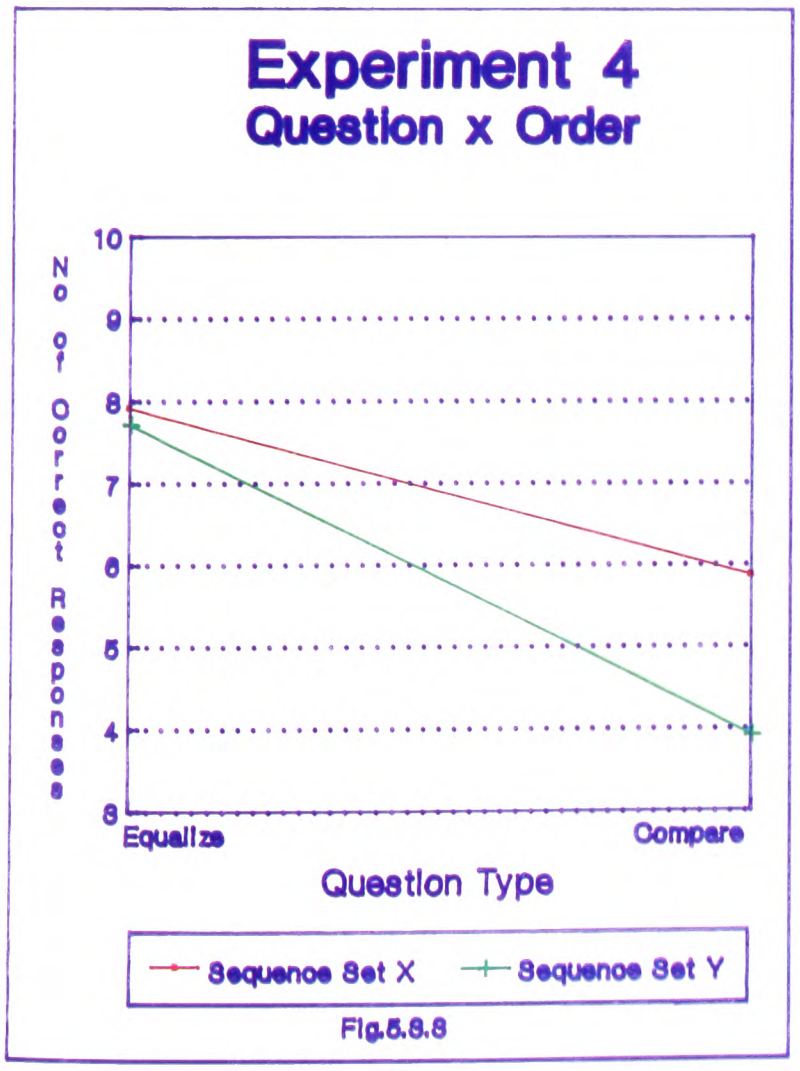


Fig.5.8.3

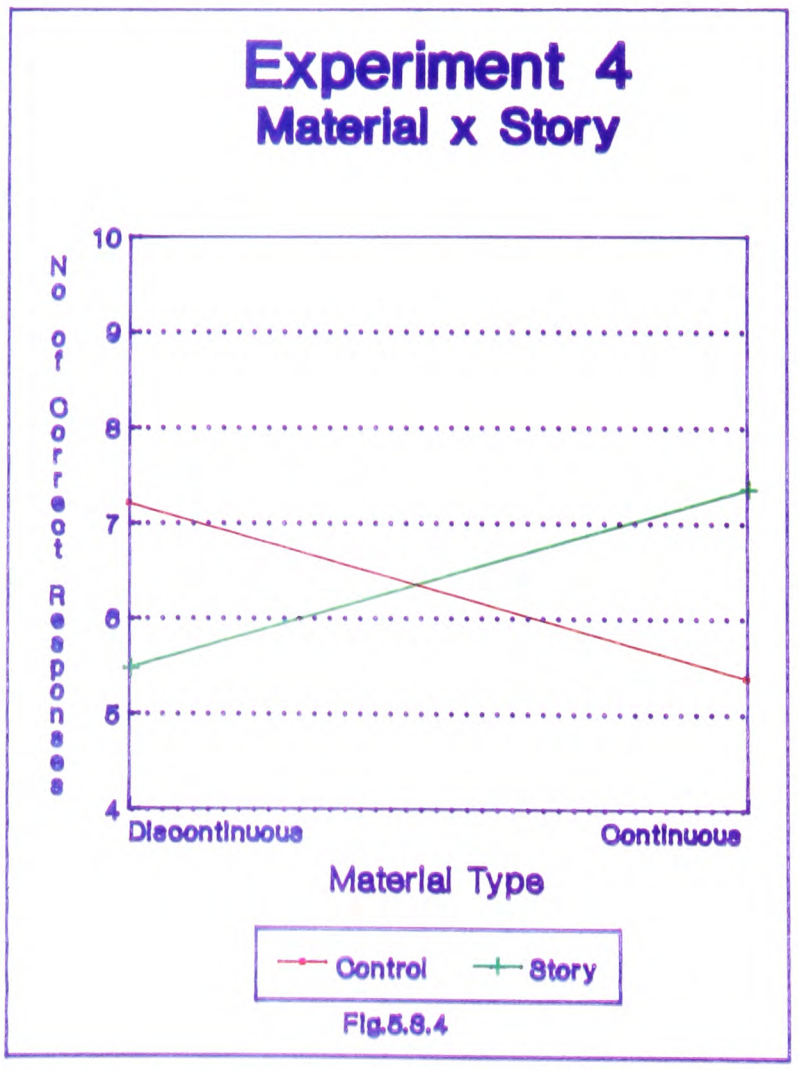


Fig.5.8.4

CHAPTER SIX

CHILDREN'S UNDERSTANDING OF CONTINUOUS GRAPHICAL INFORMATION

6 Experiment 3 An Investigation of Responses made in Different Types of Question using Different Types of Continuous Material

6.1 Introduction

Experiment 3 provided negative results with regards to the manipulation of display type. In Experiment 3, the Equalize and Compare difference was being investigated in terms of the difficulty that children have with Compare-type questions. It was thought that this difficulty may be due to a failure to understand part-whole relationships. It was hypothesized that in the previous experiments for the continuous graphical information, part of the children's difficulty in understanding part-whole relationships had to do with their difficulty in understanding that they have to cut out the larger standard into two smaller bits. However, the negative results attained in Experiment 3 did not support this hypothesis.

Experiment 5 will utilize a different tactic to investigate the Equalize and Compare difference. The Equalize and Compare difference will be investigated in terms of the relative ease that children have with Equalize-type questions. For an Equalize-type problem, the spatial imagery may involve the child's conceptualization of how far the smaller standard has to grow or of how far the larger standard has to shrink. It is quite likely that children may find the Equalize question easy, because it is easy to envisage an increase in a quantity of lesser amount by imagining such a change

through spatial imagery. When the opportunity to use spatial imagery is manipulated by presenting the choice stimuli differently from the comparative pair in a Same-type of material display and in a Different-type of material display, children should have difficulty solving the Equalize-type question. If spatial imagery and imagining the continuous stimuli growing or shrinking are accountable for the Equalize question's ease, then it should be much easier to solve this type of problem if the comparative and choice material are the same than if they are different.

A Same-type material display constituted the comparative pair and the choice stimuli to be composed of the same display [i.e. Bar-Bar and Line-Line]. A Different-type material display constituted the comparative pair and the choice stimuli to be composed of different displays [i.e. Bar-Line and Line-Bar]. The connection between the two sets should then be more difficult to make in the Equalize-type question with Different-type material, as children should find the manipulation of spatial imagery more difficult across different material. These different manipulations of continuous displays would be presented to the child to see if spatial imagery had anything to do with the Equalize and Compare difference. (See Figure 6.1.1.)

It was predicted that there would be no difference between Equalize and Compare questions with the Different-type displays because this display would make Equalize-type questions just as difficult to solve as Compare-type questions. The Equalize and Compare difference would be smaller if the material display were different than if it were the same. That is to say, the Equalize and Compare difference would be reduced in the different condition, but not in the same condition because it would be the different condition that would make the Equalize question harder.

An older age group had been omitted from Experiments 2, 3, and 4 as they had originally performed very successfully in Experiment 1, and demonstrated not to be affected by Question Type. However, as it was the

aim of this experiment to make the Equalize-type question harder by manipulating spatial imagery, an older age group was introduced in order to see the effect of this spatial manipulation on them.

6.2 Method

(a) Subjects

Two groups of children served as subjects: one consisting of 24 children ranging in age from 6;3 - 7;3 (mean age of 6;8) and the other of 24 children ranging in age from 7;5 - 8;11 (mean age of 8;1). Six additional children, ranging in age from 5;4 - 6;4 (mean age of 5;8), served as pilot subjects. The children were all attending a first school in Oxford, England.

(b) Apparatus

The same apparatus was used as for Experiments 2, 3, and 4 with the exception that an Apple II microcomputer was utilized instead of a BBC Master microcomputer.

(c) Design

A mixed design with repeated measures on Conditions (3: Material [M], Condition [C], and Question Type [Q]) was used. Age (2: 6, 8) was the Between factor.

Each child was given 64 trials, over two sessions on two separate days (there was an interval of 10 days between one session and the next). There were 64 trials in each sequence. One session of 32 trials was for the Equalize-type questions and the other session of 32 trials was for the

Compare-type questions. One session of 32 trials was also for the Same-type of material display and the other session of 32 trials was also for the Different-type of material display. (As mentioned earlier, a Same-type material display constituted the comparative pair and the choice stimuli to be composed of the same display [i.e. Bar-Bar and Line-Line] and a Different-type material display constituted the comparative pair and the choice stimuli to be composed of different displays [i.e. Bar-Line and Line-Bar].) Each of the 32 trials were blocked into eight different randomized sequences of the displays. There were four counterbalanced orders of the eight blocked displays. Four groups of six children each received one of these orders. The four orders were the following.:

Order 1

First session

Display	Question Type
BB	1
	2
BB	2
LL	1

Second session

Display	Question Type	
BL	1	BL
LB	2	
LL	2	
LB	1	

Order 2

First session

Display	Question Type
BL	1
BB	2
BL	2
LL	1

Second session

Display	Question Type
LB	1
LL	2
LB	2
BB	1

Order 3

First session

Display	Question Type
LL	1
LB	2
LL	2
BL	1

Second session

Display	Question Type
BB	1
BL	2
BB	2
LB	1

Order 4

First session

Display	Question Type
LL	2
LB	1
BB	1
LB	2

Second session

Display	Question Type
BL	2
LL	1
BL	1
BB	2

where the material displays were: (see Figure 6.1.1):

(1) Condition BB. The comparative pair and the choice stimuli were bars. This was a within (Same-material) condition.

(2) Condition BL. The comparative pair were bars and the choice stimuli were lines. This was a between (Different-material) condition.

(3) Condition LL. The comparative pair and the choice stimuli were lines. This was a within (Same-material) condition.

(4) Condition LB. The comparative pair were lines and the choice stimuli were bars. This was a between (Different-material) condition.

and where the question type presentations were:

1 = Equalize-type questions

2 = Compare-type questions

The question types were presented in alternating order. If in one display, children received the Equalize-type question first, in the other display they received the Compare-type question first. The design for Experiment 5 is illustrated in Table 6.2.1.

(d) Materials

The continuous computer-driven displays used are presented in Figure 6.1.1.

I. Same-type Material Display

This type of display constituted the comparative pair and the choice stimuli to be composed of the same display.

A. Bar-Bar Display

Comparative Pair

On the screen, seven stimuli were presented. (See Figure 6.1.1.) Depending on the condition, two of these stimuli, which appeared on the left-hand side of the display, and which were separated from the other five stimuli, consisted of either of (a) two bars or of (b) two lines. These two stimuli were the comparative pair, as previously described in Experiments 2, 3, and 4.

This comparative pair was taken randomly from a set of 11 bars which were 1,2,3,4,5,6,7,8,9,10, and 11 inches high. In four of the displays, the bar on the far left was the taller, labelled as the larger standard. It was either 11,10,7, or 6 inches high. The column of bars next to it was the shorter, labelled as the smaller standard. It was either 2 or 3 inches high. The difference in height between the smaller standard and the larger standard was either 4 or 8 inches. In these displays, the bars were arranged in

descending order. The design of the comparative pair, when the bar on the far left was the larger standard, was as follows:

(larger standard) inches in height	4	8 (difference in height)
11	-	3
10	-	2
7	3	-
6	2	-
	(smaller standard) inches in height	

In the other four displays, the bar on the far left was the shorter, labelled as the smaller standard. It was either 2 or 3 inches high. The column of bars next to it was the taller, labelled as the larger standard. It was either 4,5,8, or 9 inches high. The difference in height between the smaller standard and the larger standard was either 2 or 6 inches. In these displays, the bars were arranged in ascending order. The design of the comparative pair, when the bar on the far left was the smaller standard, was as follows:

(smaller standard) inches in height	2	6 (difference in height)
2	4	-
3	5	-
2	-	8
3	-	9
	(larger standard) inches in height	

Choice Stimuli

The remaining five stimuli, which appeared on the right-hand side of the display, consisted of five bars. These five stimuli were the choice stimuli, as the child had to choose the one which was representative of the difference between the comparative pair.

These choice stimuli were taken from a set of bars which were 2,4,6,8, and 10 inches high. In half the displays, these bars were arranged in

ascending order (2,4,6,8, and 10) and in the other half they were arranged in descending order (10,8,6,4, and 2). In only half of the displays were the choice stimuli of the same height as the comparative stimuli.

As mentioned in the design section (c), there were eight fixed trials. The choice stimuli that were arranged in ascending order were as follows:

Comparative Pair	Choice Stimuli
2-4 inches in height	2,4,6,8, and 10 inches in height
7-3 inches in height	2,4,6,8, and 10 inches in height
2-8 inches in height	2,4,6,8, and 10 inches in height
11-3 inches in height	2,4,6,8, and 10 inches in height

The choice stimuli that were arranged in descending order were as follows:

Comparative Pair	Choice Stimuli
3-5 inches in height	10,8,6,4, and 2 inches in height
6-2 inches in height	10,8,6,4, and 2 inches in height
3-9 inches in height	10,8,6,4, and 2 inches in height
10-2 inches in height	10,8,6,4, and 2 inches in height

The occurrence of the correct response would appear at least one time in each of the five choice stimuli. Thus, the correct response was in the first position in one of the displays; in the second position in two of the displays; in the third position in two of the displays; in the fourth position in two of the displays; and in the fifth position in one of the displays.

In all displays, the items (bars) in each set were positioned one inch apart, and there was a distance of 2 inches between the two stimuli sets.

Any display could be green, pink, blue, or yellow in colour, and the colour of any given display was selected randomly.

B. Line-Line Displays

The Line-Line Displays were similar to the Bar-Bar displays, except that the quantities were here represented by lines. The same design was

used as for the Bar-Bar displays, except that, for example, a 7-inch bar was replaced by a 7-inch line. (See Figure 6.1.1.)

II. Different-type Material Display

C. Bar-Line Displays

The Bar-Line Displays were similar to the Bar-Bar and Line-Line displays, except that the quantities were represented by both bars and lines. The same design was used as for the Bar-Bar and Line-Line displays, except that, the comparative pair was here represented by bars and the choice stimuli by lines. (See Figure 6.1.1.)

D. Line-Bar Displays

The Line-Bar Displays were similar to the Bar-Bar and Line-Line displays, except that the quantities were represented by both bars and lines. The same design was used as for the Bar-Bar and Line-Line displays, except that, the comparative pair was here represented by lines and the choice stimuli by bars. (See Figure 6.1.1.)

The difference between Experiment 5 from Experiments 2, 3, and 4 can be summarized in the following way. The displays included (1) a Same-type material display constituting the comparative and choice stimuli to be composed of the same display (either BB displays or LL displays), and (2) a Different-type material display constituting the comparative pair and the choice stimuli to be composed of different displays (either BL displays or LB displays). (See Figure 6.1.1.)

(e) **Procedure**

Children were taken from their classes individually, in a random order. The children were shown each display one at a time as presented on the screen. Each of the eight displays was presented twice under the question-type condition ("Equalize" or "Compare").

For each display, children were asked two questions (eight questions in the Equalize form and eight questions in the Compare form). Hence, each child was asked 16 questions per display (Display BB, Display BL, Display LL, or Display LB), totalling 64 questions in all.

The experimenter asked the child four Equalize-type questions of the form, *"How much more does this one* (pointing to the smaller standard) *need to have, to have the same as this one* (pointing to the larger standard)*?"*. The four Compare-type questions were of the form, *"How much more does this one* (pointing to the larger standard) *have than this one* (pointing to the smaller standard)*?"*. The other four Equalize-type questions were of the form, *"How much less does this one* (pointing to the smaller standard) *need to have, to have the same as this one* (pointing to the larger standard)*?"*. The other four Compare-type questions were of the form, *"How much less does this one* (pointing to the smaller standard) *have than this one* (pointing to the larger standard)*?"*. The child then selected an answer (one of the five bars or lines of the choice stimuli).

As in the previous experiments, responses were recorded and stored in the computer.

6.3 Results

The principal data were the number of questions answered correctly. The mean number of correct responses for each type of question is shown in Table 6.3.1. These means were obtained from scores on the 8 questions on

each of the four types of display and give an overall picture of children's performance.

It was predicted that there would be no difference between Equalize and Compare questions in the different condition as the display manipulations would make the Equalize questions just as difficult to solve as the Compare questions, but there would be a difference in the same condition. Hence, a Question x Condition interaction was expected, where the Equalize and Compare difference would come up in the same condition, but not in the different condition.

However, in general, once again and contrary to the prediction, more correct responses occurred in the Equalize questions than in the Compare questions. Overall, there did not seem to be any difference in the amount of correct responses to Equalize and Compare questions in either the same- or different-type conditions. It seems that display manipulation does not affect the Equalize and Compare difference.

Table 6.3.1 **Means and Standard Deviations (out of 8) for the Number of Correct Responses**

6.3.1 Comparing Correct Scores across Conditions

The means and standard deviations for the total number of correct responses made in each display by the children are given in Table 6.3.1. This table shows that most correct responses were made by the older age group than by the younger age group. Contrary to prediction, Equalize questions were superior to Compare questions. Same-type of material display yielded more correct responses from the children than different-type of material display. That is to say, children performed better in the BB and LL conditions than in the BL and LB conditions.

Calculating the Probability of Correct Responses occurring purely by Chance

As for Experiments 2 and 3, children had a choice of five responses (from the choice stimuli). Therefore, there was a 20% chance of their choosing any one of these responses. Children were considered to be performing at chance level if they got a total of 1.6 correct responses. The probability that the correct responses occurred simply by chance was explored using a modification of the standard hypothesis testing procedure with the means in Table 6.3.1. (See Section 3.3.1.)

The proportions of correct responses in all means were significantly ($p<0.01$) greater than expected by chance. This indicates that the children were not acting randomly.

6.3.2 Analysis of Variance

These trends were analyzed by subjecting the raw scores for the total number of correct responses to a 2x2x2x2 ANOVA of mixed design. There was one between-subject variable: age-group (younger or older). There were three within-subject variables: Material in Comparative Pair (Bars or Lines), Condition (Comparative Pair and Choice Stimuli Same Material or Comparative Pair and Choice Stimuli Different Material), and Question Type (Equalize or Compare). The results for this analysis are presented in Table 6.3.2.

The significant main effect of Condition ($F(1,42)=15.59$, $p<0.001$) indicated that children performed better when same-type displays were presented than when different-type displays were presented. Children were better in the BB and LL Conditions than the BL and LB Conditions.

The significant main effect of Question ($F(1,42)=8.36$, $p=0.01$) indicated that Equalize questions were once again easier than Compare questions.

The significant main effect of Age-group ($F(1,42)=7.94$, $p=0.01$) indicated older children performed significantly better than younger children.

There were no interaction effects and no other significant main effects. The expected Question x Condition interaction did not occur. This indicates that the manipulation of display did not affect the Equalize or Compare questions.

6.4 Discussion

The results of Experiment 5 indicated that the older children (mean age of 8;1) performed better than the younger children (mean age of 6;8). Results confirmed children's difficulty with Compare-type questions and their relative ease with Equalize-type questions. Furthermore, results also indicated an evident display type effect. Children performed better when the two sets of items were of the same material than when they were of different materials. However, results further indicated that question type and display types are independent from each other.

It was predicted that a Question by Condition interaction would occur. Children were expected to find the Equalize question with different material more difficult, hence reducing their relative ease in performance. This interaction was expected as there should not have been any difference between Equalize and Compare questions under the different-type condition, as the Equalize and Compare difference was expected to come up in the same condition, but not in the different condition. However, the overall effects demonstrated that the expected Question by Condition interaction did not occur. The Different condition was just generally harder than the same

condition and the Equalize and Compare difference applies to both same and different material, but they do not interact.

This experiment was meant to demonstrate that the effects of spatial imagery would make the Equalize question as difficult as the Compare question. This did not happen. Though children cannot perform as well on the Compare-type word problems as on the Equalize-type word problems nor on different material displays as on same material displays, performance on both these type word problems was not affected by manipulations of material (that is, using the same-type continuous material, or a different-type continuous material). There was no sign of the predicted interaction, which would have indicated the child's understanding of spatial imagery.

The Equalize and Compare difference is strongly supported by the results of the present experiment. The question independent effect clearly demonstrates children's ease in solving the Equalize-type questions and their difficulty in solving the Compare-type questions. The display independent effect clearly demonstrates that children prefer same-type displays than different-type displays. However, these do not seem to affect children's performance on question type. Hence, Experiment 6 was then designed to pursue the spatial imagery hypothesis by introducing different level displays.

FIGURES AND TABLES FOR EXPERIMENT 5

Experiment 5
Example of Bar-Bar Display

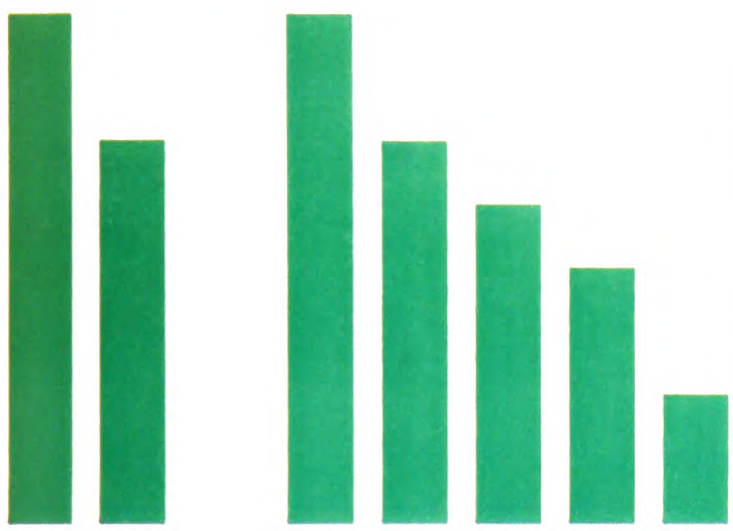


Fig.6.1.1

Experiment 5
Example of Line-Line Display

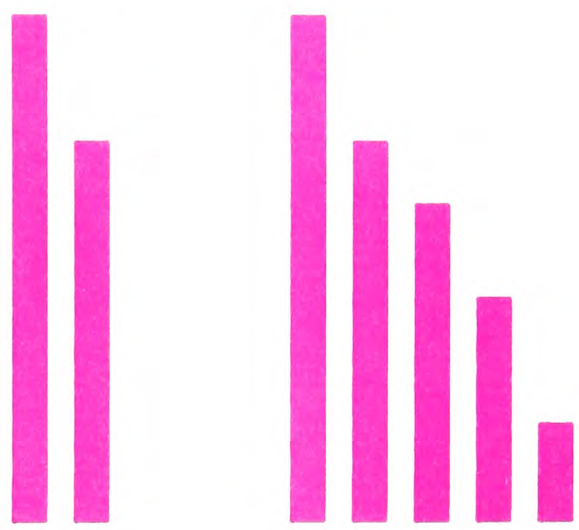


Fig.6.1.1

Experiment 5
Example of Bar-Line Display

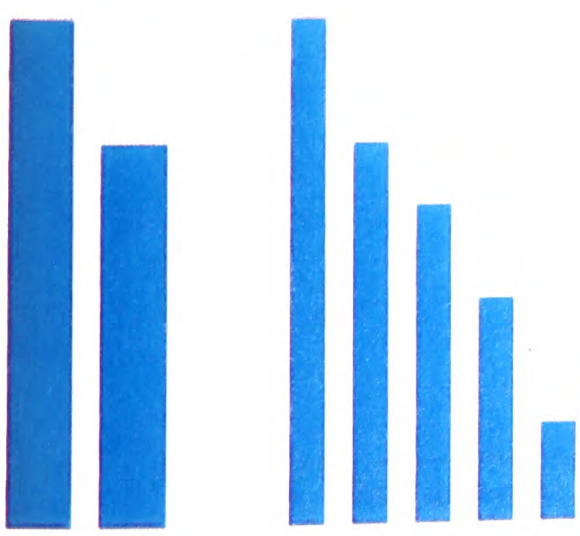


Fig.6.1.1

Experiment 5
Example of Line-Bar Display

Fig.6.1.1

Table 6.2.1 The Design of Experiment 5

No. of children	No. of children in each sequence	Material Order	Material Display	Question Type (Sequence Set)
(Young Group) 24	6	1	BB	x
			BL	y
			LL	x
			LB	y
	6	2	BB	y
			BL	x
			LL	x
			LB	x
	6	3	BB	x
			BL	x
			LL	x
			LB	y
	6	4	BB	x
			BL	y
			LL	y
			LB	x
(Old Group) 24	6	1	BB	x
			BL	y
			LL	x
			LB	y
	6	2	BB	y
			BL	x
			LL	x
			LB	x
	6	3	BB	x
			BL	x
			LL	x
			LB	y
	6	4	BB	x
			BL	y
			LL	y
			LB	x

x ▪ Sequence Set Equalize first, Compare second
y ▪ Sequence Set Compare first, Equalize second

Table 6.3.1 Means and Standard Deviations (out of 8) for the Number of Correct Responses

Age Group	Material Group		Question Type: Equalize	Question Type: Compare
Young	SAME	BB	4.64 (3.26)	3.86 (3.20)
		LL	4.64 (2.84)	4.23 (3.09)
	DIFFERENT	BL	4.18 (2.97)	3.55 (3.13)
		LB	4.82 (2.99)	3.68 (3.06)
Old	SAME	BB	6.73 (2.45)	6.50 (2.45)
		LL	6.73 (2.43)	6.46 (2.46)
	DIFFERENT	BL	6.36 (2.50)	6.05 (2.44)
		LB	6.18 (2.28)	6.14 (2.59)

(Standard deviations are given in brackets.)

Table 6.3.2 Summary Table (ANOVA) for Total Correct Scores

Source	SS	DF	MS	F	PROB.
Mean	9870.727	1	9870.727	185.10	0.0000
AgeGroup[A]	423.284	1	423.284	7.94	0.0073
error	2239.739	42	53.327		
Question[Q]	20.046	1	20.046	8.36	0.0060
QA	6.011	1	6.011	2.51	0.1208
error	100.693	42	2.398		
Material[M]	1.375	1	1.375	1.46	0.2342
MA	2.227	1	2.227	2.36	0.1320
error	39.648	42	0.944		
QM	0.011	1	0.011	0.02	0.8887
QMA	0.182	1	0.182	0.32	0.5762
error	24.057	42	0.573		
Condition[C]	10.921	1	10.921	15.59	0.0003
CA	0.409	1	0.409	0.58	0.4490
error	29.421	42	0.701		
QC	0.284	1	0.284	0.25	0.6197
QCA	0.727	1	0.727	0.64	0.4283
error	47.739	42	1.137		
MC	0.182	1	0.182	0.15	0.6973
MCA	0.284	1	0.284	0.24	0.6270
error	49.784	42	1.185		
QMC	0.409	1	0.409	0.29	0.5951
QMCA	1.921	1	1.921	1.35	0.2525
error	59.921	42	1.427		

SS = Sum of Squared Deviations
DF = Degrees of Freedom
MS = Mean of Squared Deviations
F = Ratio of Variances
Prob. = Level of Significance

CHAPTER SEVEN

CHILDREN'S UNDERSTANDING OF CONTINUOUS GRAPHICAL INFORMATION IN DIFFERENT LEVEL DISPLAYS

7 Experiment 6 An Investigation of Responses made in Different Types of Question using Different Types of Displays with Continuous Material

7.1 Introduction

Experiment 5 had suggested that perceptual factors play some part in the understanding of both Equalize and Compare questions. This was demonstrated by the effect of condition where same-type material displays were easier than different-type material displays. However, there was no interaction between Question Type and Condition, and so the perceptual effects did not explain the Equalize and Compare difference, because they applied to both question types.

One form of Riley, Greeno, and Heller's action hypothesis is that children might be using some kind of spatial imagery and spatial manipulation to help solve the Equalize problems. They may be imagining each choice stimulus being added to the smaller standard of the comparative pair to see if that combination equals the larger standard. By the same token, they may be imagining each choice stimulus being taken away from the larger standard of the comparative pair to see if that subtraction equals the smaller standard. In that case, one would have to consider the nature of the spatial manipulation. This would actually be quite a complex manipulation, since it would involve moving a spatial image of each choice stimulus upwards and to the left in order to add it to the smaller stimulus (or subtract it from the larger stimulus)

of the comparative pair. In this case, the complex movement may itself be a problem for young children and it follows that making the necessary movement a simpler one might affect the success of children's judgements.

This can be done by setting up what this author calls "different level displays", in which the choice stimuli have a higher baseline than the comparative pair and that baseline is on a same level with the top of the smaller stimulus of the comparative pair. This means that the imagined spatial movement of the images of the choice stimuli would only be a lateral one. (See Figure 7.1.1.)

Brandimonte, Hitch, and Bishop (1992) discussed how visual patterns can be transformed and reinterpreted in mental imagery. They tested children and adults in two tasks: a combination task, where images could be combined, and a subtraction task, where parts of the image could be removed so as to form new images. Children of 6-to-10-years were able to transform a mental image, as well as reinterpret it, so as to form another image with another interpretation. Their study contradicts the mental imagery theory of Piaget and Inhelder (1971) who claimed that young children are only capable of static imagery and are not able to transform images until 7-to-8-years of age.

With regard to this study, this author's view of spatial imagery manipulation is similar to Brandimonte et al.'s analysis of the combination task. Their combination task entailed superimposing two different images in order to discover a new image. In the case of this experiment, the manipulation involves superimposing two different and similar images in order to discover a new image. Brandimonte et al.'s subtraction task entailed mentally taking away a part from an image in order to discover a new image. In the case of this experiment, the manipulation for the comparative term "less" involves the same concept as Brandimonte et al.'s subtraction task.

Brandimonte et al. concluded that children are able to transform visual mental images into new images. Hence, children can manipulate mental images quite successfully. Based on these results, it is thought that children should have no difficulty with this experiment's spatial imagery manipulation.

The idea of this spatial imagery move is further supported by the observed strategies in Experiments 2 (see Section 3.3.2), 3 (see Section 4.3.2), and 4 (see Section 5.3.2). Children used the Measuring Strategy with the Equalize question more frequently than with the Compare question. This Measuring Strategy, which entailed making an approximate measurement of the empty space from the smaller addend to the larger addend, may be an observed representation of the child's making of this spatial imagery move.

When the comparative pair is on a different level from the choice stimuli, there are three possibilities that could arise with regard to the effect that this manipulation has on the Equalize and Compare difference in Experiment 6. These are:

(1) The Equalize and Compare difference may diminish as this spatial imagery move may make Compare problems easier to solve.

(2) The opposite effect of (1) may happen. The above-mentioned spatial imagery move may be structurally inherent in the Equalize question. Children may be automatically cued to use this spatial imagery manipulation in the Equalize-type questions. Hence, Equalize questions may be helped by having the material at a different level. A same level display would constitute the comparative and choice stimuli spatially arranged to begin at the same base level. A different level display would constitute the comparative and choice stimuli spatially arranged to begin at different base levels.

(3) Both Equalize- and Compare-type questions will be helped.

Two pilot experiments (Appendices 4 and 5) provided evidence that varying the level of the displays has an effect on tasks involving length comparison. This makes it possible for same level and different level displays

to have an effect on children's performance in Equalize and Compare questions.

In previous experiments there was some sign that children used the Equivalence Strategy with the Compare question more frequently than with the Equalize question. Though not a consistent indication, there were definite signs of this particular strategy use in Experiment 4. This Equivalence Strategy entailed selecting a choice stimulus equivalent to one of the comparative pair. This Equivalence Strategy was further confirmed upon a breakdown of children's errors. Most of the errors in the Compare-type question entailed selecting the choice stimulus that was the same absolute size as one of the comparative pair. It seems that in the Compare questions, children may be forced back to an original value of one of the comparative stimuli. This leads to the prediction that children should be helped in the Compare question, if none of the choice stimuli is the same absolute size as the stimuli in the comparative pair. It is hypothesized that children should do better on Compare questions in Experiment 6, as the comparative pair will not be among the choice stimuli.

Experiment 6 will investigate children's ability to represent the difference between two quantities via Equalize and Compare questions with same and different material in same and different level displays.

7.2 Method

(a) Subjects

Two groups of children served as subjects: one consisting of 20 children ranging in age from 5;3 - 6;6 (mean age of 6;1), and the other of 20 children ranging in age from 6;7 - 7;7 (mean age of 7;1). The children were all attending a first school in Oxford, England.

(b) **Apparatus**

The same apparatus was used as for Experiments 2, 3, and 4.

(c) **Design**

A mixed design with repeated measures on Conditions (5: Question Type [Q], Comparative Term [CT], Display Type [D], Condition [C], and Material [M]) was used. Age (2: 6, 7) was the Between factor.

Each child was given 96 trials, over two sessions on two separate days (there was an interval of 14 days between one session and the next). One session of 48 trials was for the "more" comparative term and the other session of 48 trials was for the "less" comparative term. Twenty-four of these 48 trials were for the Equalize questions and the other 24 trials were for the Compare questions. Twelve of these 24 trials were for the same level display and the other 12 trials were for the different level display.

A same level display constituted the comparative pair and choice stimuli spatially arranged to begin at the same base level. A different level display constituted the comparative pair and choice stimuli spatially arranged to begin at different base levels; in this spatial arrangement the choice stimuli was higher than the comparative pair. The base level of the choice stimuli was raised to the height of the shortest item of the comparative pair. (See Figure 7.1.1.) There were 48 trials and each had a different presentation. Each of the 48 trials was blocked into four conditions of six displays each (three same level displays and three different level displays). There were four counterbalanced orders of the four conditions. Four groups of 5 children each received one of these orders. The four orders were the following:

Order 1

First session

Display	Comparative Term
BB	1
LL	1
BL	1
LB	1

Second session

Display	Comparative Term
BB	2
LL	2
BL	2
LB	2

Order 2

First session

Display	Comparative Term
LL	2
BL	2
LB	2
BB	2

Second session

Display	Comparative Term
LL	1
BL	1
LB	1
BB	1

Order 3

First session

Display	Comparative Term
BL	2
LB	2
BB	2
LL	2

Second session

Display	Comparative Term
BL	1
LB	1
BB	1
LL	1

Order 4

First session		Second session	
Display	**Comparative Term**	**Display**	**Comparative Term**
LB	1	LB	2
BB	1	BB	2
LL	1	LL	2
BL	1	BL	2

where for Comparative Term:

1 = "More" comparative term

2 = "Less" comparative term

where the material displays were:

(1) Condition BB. The comparative pair and the choice stimuli were bars. This was a within (same material) condition.

(2) Condition BL. The comparative pair was a bar and the choice stimuli were lines. This was a between (different material) condition.

(3) Condition LL. The comparative pair and the choice stimuli were lines. This was a within (same material) condition.

(4) Condition LB. The comparative pair was a line and the choice stimuli were bars. This was a between (different material) condition.

The question types were also presented in alternating order. Half of the children in one condition received the Equalize-type question first followed by the Compare-type question (sequence-set X); the other half were given the task in the reverse order (sequence-set Y).

The display variable was presented in alternating order. For both groups of children, half of the subjects received the same level display first followed by the different level display; the other half were given the task in the reverse order. The design for Experiment 6 is illustrated in Table 7.2.1.

(d) Materials

The continuous computer-driven displays for both same level and different level displays are presented in Figure 7.1.1.

I. Same-type Material Display

This type of display constituted the comparative pair and the choice stimuli to be composed of the same display.

A. Bar-Bar Display

Comparative Stimuli

On the screen, five stimuli were presented. The comparative pair consisted of two bars. They were taken randomly from a set of 8 bars, which were 1, 2, 3, 4, 5, 6, 7, and 8 inches high. The bar on the far left was always the taller, labelled as the larger standard. It was always 8, 7, or 6 inches in height. The bar next to it was always the shorter, labelled as the smaller standard. It was always either 5, 4, 3, 2, or 1 inch high. The difference in height between the smaller standard and the larger standard was always either 3, 4, or 5 inches. In these displays, the bars were arranged in descending order. Hence, the design of the comparative pair was as follows:

(larger standard) inches in height	3	4	5	(difference)
8	5	4	3	
7	4	3	2	
6	3	2	1	
	(smaller standard) inches in height			

Choice Stimuli

The remaining three stimuli, which appeared on the right-hand side of the display, consisted of three bars, which were 1, 2, 3, 4, and 5 inches high. The choice stimuli was taken from the design for the smaller standard of the comparative pair, as shown above. In half the displays, these bars or lines were arranged in ascending order (3,4, and 5 inches in height; 2,3, and 4 inches in height; and 1,2, and 3 inches in height) and in the other half they were arranged in descending order (5,4, and 3 inches in height; 4,3, and 2 inches in height; and 3,2, and 1 inch in height). The choice stimuli was not a subset of the comparative pair.

Each of the 48 trials had a different presentation and was blocked into two sets of 24 trials each. One set was for the Equalize question and the other for the Compare questions. Each set was blocked into four conditions of six displays each (three same level displays and three different level displays). Each different display had a trial label, represented by a number. As mentioned in the design section (c), there were 48 fixed trials. These were arranged as follows:

Equalize Questions

Condition	Trial	AA	BB	V	CR	DR
	1.	1	2	1	3	1
	2.	2	3	2	1	2
BB	3.	3	1	1	2	2
	4.	1	3	1	1	2
	5.	2	1	2	3	1
	6.	3	2	2	2	1
	7.	2	1	2	3	2
	8.	1	3	1	2	2
LL	9.	3	2	1	1	1
	10.	1	3	2	1	1
	11.	2	1	1	3	2
	12.	3	2	2	2	1
	13.	1	2	1	3	2
	14.	2	3	1	1	1
BL	15.	3	1	2	2	2
	16.	2	2	2	3	1
	17.	3	1	1	2	1
	18.	1	3	2	1	2
	19.	1	2	2	3	2
	20.	2	3	2	1	1
LB	21.	3	1	2	2	1
	22.	3	1	1	1	1
	23.	1	3	1	2	2
	24.	2	2	1	3	2

Compare Questions

Condition	Trial	AA	BB	V	CR	DR
	25.	1	2	2	3	2
	26.	2	3	2	1	1
BB	27.	3	1	2	2	2
	28.	1	1	1	3	1
	29.	3	2	1	1	2
	30.	2	3	1	2	1
	31.	2	1	1	3	1
	32.	1	3	2	2	2
LL	33.	3	2	1	1	2
	34.	1	1	2	3	2
	35.	3	2	2	1	1
	36.	2	3	1	2	1
	37.	1	2	2	3	1
	38.	2	3	1	1	2
BL	39.	3	1	2	2	1
	40.	3	2	1	2	2
	41.	2	3	2	1	2
	42.	1	1	1	3	1
	43.	2	1	1	3	1
	44.	1	3	1	2	1
LB	45.	3	2	1	1	2
	46.	3	2	2	2	2
	47.	2	3	2	1	1
	48.	1	1	2	3	2

AA=Height of the larger standard,where
1=8 inches high;
2=7 inches high; and
3=6 inches high.

BB=Difference between the two standards, where
1=the difference of 3 inches;
2=the difference of 4 inches; and
3=the difference of 5 inches.

V= Condition type, where
1=standing condition, and
2=raised condition.

CR=Location of correct response, where
1=correct choice is the smaller of the choice stimuli;
2=correct choice is the middle of the choice stimuli; and
3=correct choice is the largest of the choice stimuli.

DR=Ascending or descending choice of displays, where
1=ascending display, and
2=descending display.

The occurrence of the correct response would appear at least one time in each of the three choice stimuli. Thus, the correct response was in the first position in a third of the total displays; in the second position in a third of the total displays; and in the third position in the last third of the total displays.

In all displays, the items (bars) in each set were positioned one inch apart, and there was a distance of 2 inches between the two stimuli sets.

Any display could be green, pink, blue, or yellow in colour, and the colour of any given display was selected randomly. (See Figure 7.1.1.)

B. Line-Line Displays

The Line-Line Displays were similar to the Bar-Bar displays, except that the quantities were represented by lines. The same design was used as for the Bar-Bar displays, except that, for example, a 7-inch bar was replaced by a 7-inch line. (See Figure 7.1.1.)

II. Different-type Material Display

C. Bar-Line Displays

The Bar-Line Displays were similar to the Bar-Bar and Line-Line displays, except that the quantities were represented by both bars and lines. The same design was used as for the Bar-Bar and Line-Line displays, except that, the comparative pair was here represented by bars and the choice stimuli by lines. (See Figure 7.1.1.)

D. Line-Bar Displays

The Line-Bar Displays were similar to the Bar-Bar and Line-Line displays, except that the quantities were represented by both bars and lines. The same design was used as for the Bar-Bar and Line-Line displays, except that, the comparative pair was here represented by lines and the choice stimuli by bars. (See Figure 7.1.1.)

(e) Procedure

Children were taken from their classes individually, in a random order. The children were shown each display one at a time as presented on the screen. Each of the 48 displays was presented twice under the comparative term condition ("more" or "less").

For each display, children were asked two questions (48 questions in the "more" comparative term and 48 questions in the "less" comparative

term). Twenty four of these questions were in the Equalize form and 24 questions were in the Compare form. Hence, each child was asked a total of 96 questions in all.

The phrasing of Equalize-type question was of the form, *"How much more does this one* (pointing to the smaller standard) *need to have, to have the same as this one* (pointing to the larger standard)*?"*. The phrasing of a Compare-type question for the same display was of the form, *"How much more does this one* (pointing to the larger standard) *have than this one* (pointing to the smaller standard)*?"*. The child then selected an answer (one of the three stimuli from the choice stimuli).

In the other comparative term version of the task, a child was asked a further 24 questions per question type, totalling 48 further questions. The phrasing for an Equalize-type question was of the form, *"How much less does this one* (pointing to the larger standard) *need to have, to have the same as this one* (pointing to the smaller standard)*?"*. The phrasing of a Compare-type question for the same display was of the form, *"How much less does this one* (pointing to the smaller standard) *have than this one* (pointing to the larger standard)*?"*.

7.3 Results

The mean number of correct responses in each type of Question is presented in Table 7.3.2. The mean number of correct responses in each type of Display is presented in Table 7.3.3. The complete table presenting the mean number of correct responses for each type of condition is shown in Table 7.3.1 (see Appendix 6). Tables 7.3.2 and 7.3.3 are a breakdown of Table 7.3.1. These means were obtained from scores on the 48 questions on each of the four types of material display for each comparative term and give an overall picture of children's performance.

It was predicted that children would find different level displays particularly easier when presented with the same-type material for the Equalize question.

In general, there did not seem to be an Equalize and Compare difference either in the same level displays or in the different level displays, in the same or in the different material. Contrary to prediction, the Question x Display x Condition interaction did not seem to be significant.

Table 7.3.1 Means and Standard Deviations (out of 3) for the Number of Correct Responses

Table 7.3.2 Means and Standard Deviations (out of 3) for the Number of Correct Responses in each Question Type

Table 7.3.3 Means and Standard Deviations (out of 3) for the Number of Correct Responses in each Display Type

7.3.1 Comparing Correct Scores across Conditions

The means and standard deviations for the total number of correct responses made in each type of condition by the children were given in Table 7.3.1 (see Appendix 6) and in Tables 7.3.2 and 7.3.3. This table shows that more correct responses were made by the older group than by the younger group. Different level displays were superior to same level displays.

Calculating the Probability of Correct Responses occurring purely by Chance

In this experiment, children had a choice of three responses (from the choice stimuli). Therefore, there was a 33.3% chance of their choosing any one of these responses. Children were considered to be performing at chance level if they got a total of 1.0 correct responses. The probability that the correct responses occurred simply by chance was explored using a

modification of the standard hypothesis testing procedure with the means in Table 7.3.1. (See Section 3.3.1.)

Scores significantly above chance

The proportions of correct responses made in all, but six of the cell means, were significantly ($p<0.01$) greater than expected purely by chance. The proportions of correct responses made by three groups: (1) Equalize question, (LB Condition), in the different level display, comparative term "less", (2) Compare question, (LB Condition), in the different level display, comparative term "more", and (3) Compare question, (LB Condition), in the different level display, comparative term "less", were significantly ($p<0.05$) greater than expected purely by chance. This indicates that the children were not acting randomly.

Scores at chance

The proportion of correct responses made by three groups: (1) Equalize question, (LL Condition), in the same level display, comparative term "less", (2) Compare question, (LL Condition), in the same level display, comparative term "less", and (3) Compare question, (LB Condition), in the same level display, comparative term "less", were not significantly different from chance. This indicates that for these groups, children were acting randomly.

7.3.2 Analysis of Variance

These trends were analyzed by subjecting the raw scores for the total number of correct responses to a 2x2x2x2x2x2 ANOVA with repeated measures. There was one between-subject variable: age-group (younger: 6 or older: 7). There were five within-subject variables: (1) Question Type

[Q]:(Equalize or Compare), (2) Comparative Term [CT]:("more" or "less"), (3) Display Type [D]:(same level or different level), (4) Condition [C]:(same material = Bar-Bar [BB], Line-Line [LL] or different material = Bar-Line [BL], Line-Bar [LB]), and (5) Material [M]:(bars or lines in the comparative pair). The results of this analysis are presented in Table 7.3.4. (See Appendix 6.)

The significant main effect of Age ($F(1,38)=14.83$, $p<0.001$), indicated that the older children made more correct responses than the younger ones.

The significant main effect of Material ($F(1,38)=6.75$, $p=0.0133$), indicated that children made more correct responses when the comparative stimuli consisted of bars than when they consisted of lines.

The main effect of Display ($F(1,38)=3.91$, $p=0.055$), was a very interesting result that was close to significance, but not significant. This result may indicate a tendency for the different level display to be easier than the same level display. However, the fact that this effect failed to reach significance means that the attempt in this experiment to vary the effect of spatial representation did not work.

There was not a significant main effect of Question Type in this Experiment. This was a negative result, as well as an unexpected and new finding, as in all previous experiments there had always been a main effect of Question Type. This result indicated that there was no sign of an Equalize and Compare difference.

The significant interaction between Question x Display x Material x Age ($F(1,38)=9.97$, $p<0.01$), showed that the effect of Material (bars or lines) affects significantly the effect of Display (same level or different level) on the children 's performance on Question Type. Furthermore, this effect varied between the two age groups. This interaction was explored using a Newman-Keuls Multiple Range Test. In general, the older group was significantly better than the younger group in the Equalize questions when these were presented with a comparative pair of bars in the different level display. The

older group was superior in performance than the younger group. Hence, apart from the age difference, this four-way interaction did not show anything else and, therefore, should not be taken too seriously. A graphical representation of this comparison is presented in Figure 7.3.1.

There was a significant interaction between Question x Condition x Material ($F(1,38)=5.89$, $p<0.05$), which showed that Material (bars or lines) affects significantly the effect of Condition (same material or different material) on the children 's performance on Question Type. This interaction was explored using a Newman-Keuls Multiple Range Test. In general, children did better on Equalize questions when these were presented with a comparative pair of bars in a same-type material display. However, the post-hoc test demonstrated that each condition did not differ significantly from each other. A graphical representation of this comparison is presented in Figure 7.3.2.

There was an interaction between Condition x Age ($F(1,38)=4.07$, $p=0.051$). This result may indicate a tendency for the effect of condition to vary between the two age groups. The younger age group may not have been affected by the type of condition presented (whether it was same material of different material), whereas with the older group, same material seemed a lot easier than different material.

The three-way interaction between Question x Display x Material ($F(1,38)=4.06$, $p=0.051$), showed an almost significant tendency for the effect of Material (bars or lines) to affect Display (same level or different level) on the children's performance on Question Type. Children's performance on Equalize questions seemed to improve when these were presented with a comparative pair of bars in the different level display.

7.4 Discussion

This experiment has produced two main results, both of them negative.

(1) The effect of Display almost worked, but did not. The lack of a significant main effect of this variable indicated that there was not an effect of the spatial imagery manipulation. However, the probability was equal to 0.055, which was almost significant. Due to this fact, it is speculated that perhaps if the difference would have been greater, Display would have been significant. Hence, results are left inconclusive with regard to Display.

The failure to reach a significant main effect for Display indicated that the different-level displays were not working very well. This experiment relied heavily on the effect of display in order to attribute the Equalize and Compare difference to the action hypothesis. An effect of Display only came up in a four-way interaction.

The Question x Display x Material x Age interaction indicated that the Equalize and Compare difference in same and different level displays was only evident in some cells. The effect of same and different level displays on question type appeared significant only in conjunction with bars in the comparative pair for the older age group. Hence, the older children found Equalize questions easier when presented with a comparative pair of bars in the different level display than in the same level display. So it seems that the Equalize-type question is aided by the manipulation of a particular spatial imagery, that of different level displays, but only when the comparative pair are designated bars. However, this was the most important result as it indicated that children perform significantly better on Equalize questions with the type of spatial manipulation that involves a different level display.

The Question x Condition x Material interaction indicated that same- and different-type material (Condition) seem to affect children's performance on Equalize questions when these are presented with a comparative pair of bars. Hence, same- and different-type material affect the Equalize and

Compare difference irrelevant of Display. In general, same-type material has an effect on children's performance on the Equalize question, but the same applies to both same and different level displays.

(2) The failure to get a significant main effect of Question indicated that there was no sign of an Equalize and Compare difference. This indicated that something strange was happening, as it was the first time that this effect had disappeared. This result was quite important and surprising.

It is concluded that this was due to the fact that the comparative pair was not represented in the choice stimuli as this made the Equivalence Strategy quite impossible to use. Children were not able to use an Equivalence Strategy in this experiment. Hence, children's performance on the Compare questions was improved by not having the comparative pair represented in the choice stimuli. Children were not forced back to an original value of one of the comparative stimuli when solving Compare-type questions. So it seems that this Equivalence Strategy gets in the way of Compare questions. It is this structural aspect of the Compare problem which seems to account for its difficulty.

FIGURES AND TABLES FOR EXPERIMENT 6

Experiment 6
Bar-Bar Same Level

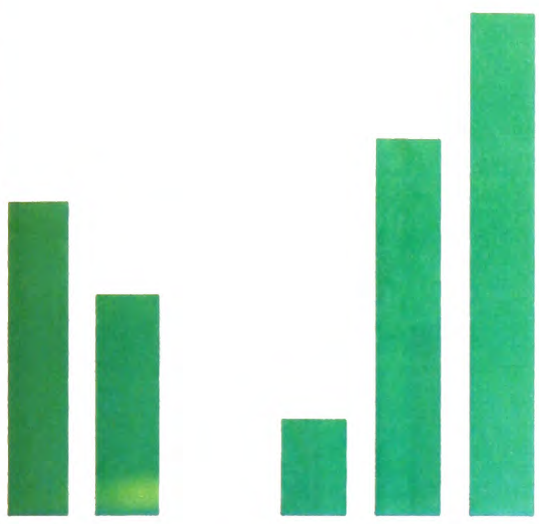


Fig.7.1.1

Experiment 6
Bar-Bar Different Level

Fig.7.1.1

Experiment 6
Line-Line Same Level

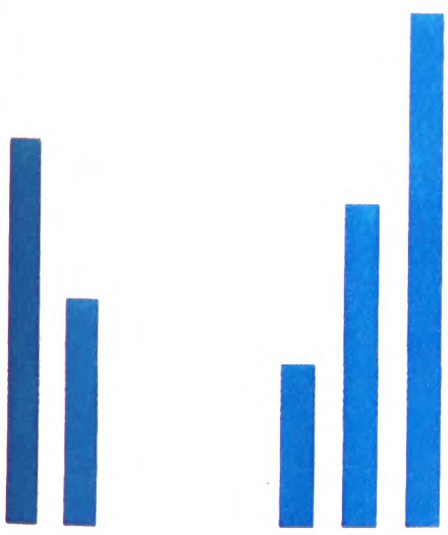


Fig.7.1.1

Experiment 6
Line-Line Different Level

Fig.7.1.1

Experiment 6
Bar-Line Same Level

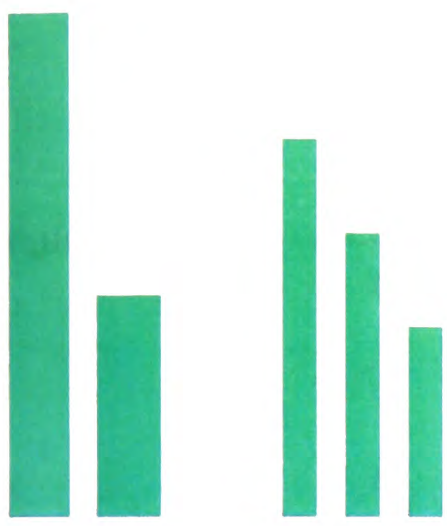


Fig.7.1.1

Experiment 6
Bar-Line Different Level

Fig.7.1.1

Experiment 6
Line-Bar Same Level

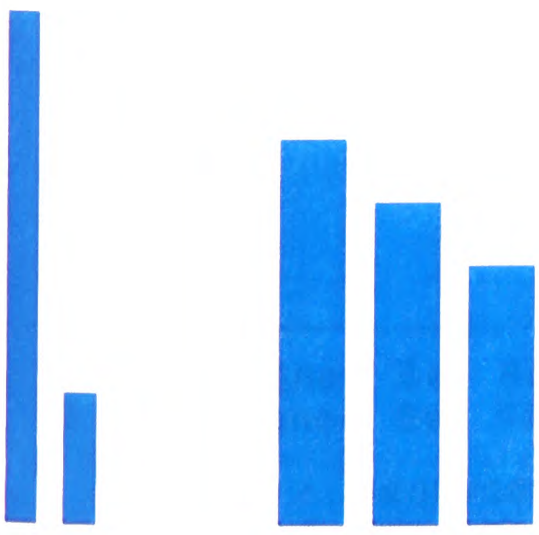


Fig.7.1.1

Experiment 6
Line-Bar Different Level

Fig.7.1.1

Table 7.2.1 The Design of Experiment 6

No. of children	Comparative Term Order Set		Display		Condition (Seq.Set)	Question (Seq.Set)
(Young Group) 20	5	1	S	BB	A	x
			A		B	y
			M	LL	B	y
			E		A	x
10	5	2	D	BL	A	x
			I		B	y
			F	LB	B	y
			F		A	x
	5	3	S	BB	B	y
			A		A	x
			M	LL	A	x
			E		B	y
10	5	4	D	BL	B	y
			I		A	x
			F	LB	A	x
			F		B	y
(Old Group) 20	5	1	S	BB	A	x
			A		B	y
			M	LL	B	y
			E		A	x
10	5	2	D	BL	A	x
			I		B	y
			F	LB	B	y
			F		A	x
	5	3	S	BB	B	y
			A		A	x
			M	LL	A	x
			E		B	y
10	5	4	D	BL	B	y
			I		A	x
			F	LB	A	x
			F		B	y

A = Sequence Set Same level first, Different Level second
B = Sequence Set Different level first, Same level second

x = Sequence Set Equalize first, Compare second
y = Sequence Set Compare first, Equalize second

Table 7.3.2 Means & Standard Deviations (out of 3) for Correct Responses in each Question Type

Age Group	Display	Comparative Term			
		More		Less	
		Equalize	Compare	Equalize	Compare
Young	BB	1.60 *0.95*	1.60 *0.98*	1.74 *0.89*	1.53 *0.90*
	LL	1.55 *0.96*	1.69 *0.88*	1.39 *0.80*	1.46 *0.96*
	BL	1.79 *0.98*	1.60 *0.91*	1.55 *1.01*	1.67 *0.96*
	LB	1.74 *0.98*	1.46 *0.81*	1.48 *1.07*	1.36 *0.89*
Old	BB	2.50 *0.77*	2.40 *0.82*	2.37 *0.91*	2.43 *0.80*
	LL	2.40 *0.73*	2.45 *0.76*	2.24 *0.92*	2.32 *0.82*
	BL	2.37 *0.76*	2.37 *0.75*	2.19 *0.91*	2.37 *0.80*
	LB	2.27 *0.83*	2.24 *0.89*	2.29 *0.87*	2.29 *0.88*

(Standard deviations in italics.)

Table 7.3.3 Means & Standard Deviations (out of 3) for Correct Responses in each Display Type

Age Group	Display	Comparative Term			
		More		Less	
		Same	Different	Same	Different
Young	BB	1.60 *1.00*	1.60 *0.92*	1.60 *0.84*	1.67 *0.95*
	LL	1.60 *0.90*	1.64 *0.94*	1.27 *0.86*	1.58 *0.89*
	BL	1.69 *0.91*	1.69 *0.99*	1.52 *0.96*	1.69 *1.01*
	LB	1.58 *0.95*	1.62 *0.83*	1.43 *0.97*	1.41 *0.99*
Old	BB	2.37 *0.82*	2.53 *0.77*	2.40 *0.83*	2.40 *0.88*
	LL	2.45 *0.72*	2.40 *0.76*	2.19 *0.91*	2.37 *0.83*
	BL	2.34 *0.75*	2.40 *0.76*	2.27 *0.86*	2.29 *0.85*
	LB	2.27 *0.81*	2.24 *0.91*	2.21 *0.95*	2.37 *0.80*

(Standard deviations in italics.)

Experiment 6

Question x Display x Material x Age

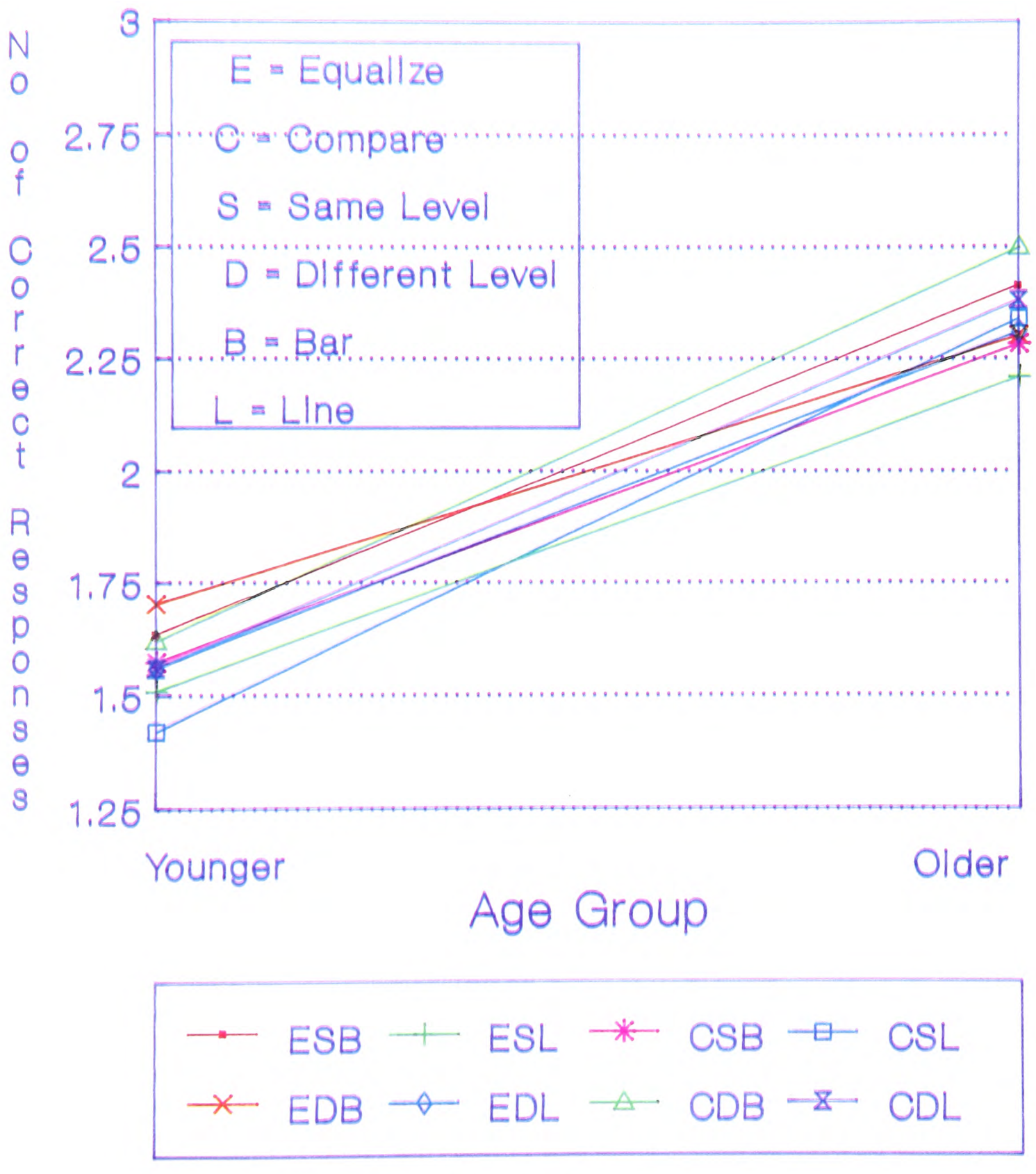


Fig.7.3.1

Experiment 6

Question x Condition x Material

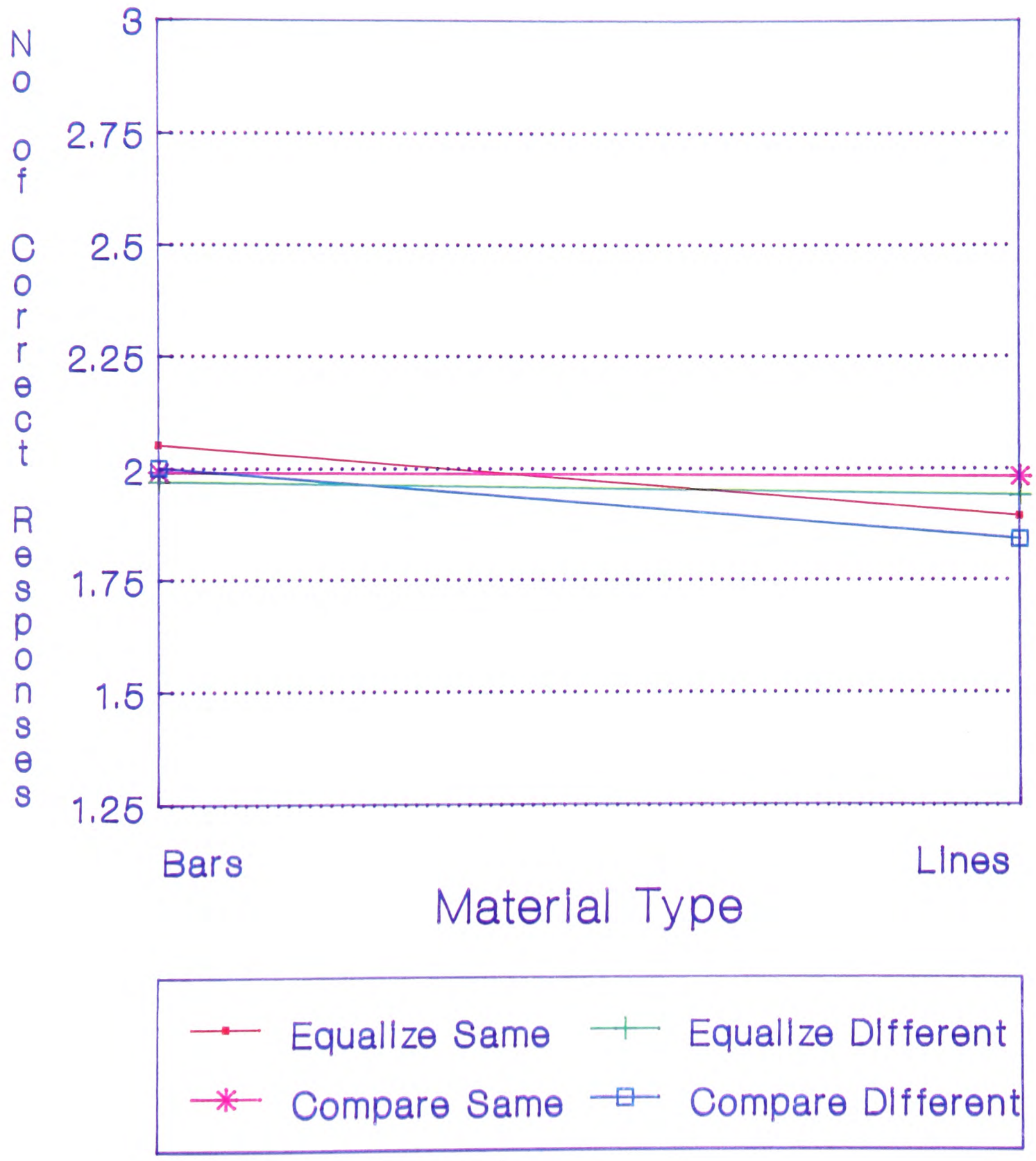


Fig.7.3.2

CHAPTER EIGHT

8 CONCLUSIONS AND RECOMMENDATIONS

8.1 Conclusions

8.1.1 A Representational Problem

This investigation concerned children's mathematical understanding and their ability to make a distinction between what a sum is on the one hand and what representing it is on the other. The best sources of information for pursuing this distinction were different types of problem which involved the same sum. Equalize and Compare word problems, as reviewed earlier (Chapter One), require the same mathematical operation; yet children seem to have difficulty with one and not so much with the other.

There is at least one reason why this distinction is of crucial importance. This leads to a matter of representation. Children prefer to solve these problems in different ways and seem to have different ways of representing them.

It was possible to look at the Equalize and Compare difference because difficulties with Compare problems had previously been reported (Riley,1981; Carpenter, Hiebert, and Moser,1981; DeCorte and Verschaffel,1987a; Hudson;1983). However, there had not been any proper systematic comparisons with Equalize word problems. Hence, there was more than one reason to investigate the comparison between Equalize and Compare word problems. These reasons were the following:

(1) To investigate whether children's mathematical understanding is fundamental to recognizing and representing the sum correctly.

(2) To compare children's difficulty with Compare word problems to their mathematical understanding of Equalize word problems.

(3) Finally, to investigate the importance of the comparison between Equalize and Compare word problems as related to how the child actually represents the problems.

8.1.2 A Quantity-Specific Problem

This thesis set out to investigate how 5-to-8-year-old children compare two unequal quantities and work out the difference between them. Of particular interest was children's performance in Equalize and Compare word problems. This thesis also set out to look at the effect on performance in Equalize and Compare problems using continuous, as well as discontinuous material.

Previous studies demonstrated children's difficulty with Compare word problems using discontinuous quantity only (Riley,1981; DeCorte and Verschaffel,1987a; Hudson,1983). The evidence was not good for the postulated Equalize and Compare difference, as there were no data providing a systematic comparison between Equalize and Compare word problems using discontinuous material. Furthermore, a comparison between Equalize and Compare word problems using continuous material had not been considered.

Researchers, such as Briars and Larkin (1984), who used discontinuous quantity in Change problems, made the underlying assumption that a difference between Equalize and Compare problems would be number-specific. Other researchers, who used discontinuous quantity in Compare problems, made the underlying assumption that a difference between Equalize and Compare problems would be quantity-specific. If quantitative comparisons are number-specific, then the Equalize and Compare difference should only be evident with discontinuous quantities (which they can, for example, count) and not with continuous quantities (which are not number related). If quantitative comparisons are non-number-specific, then the Equalize and Compare difference should be evident with both discontinuous and continuous quantities. Should the latter be the case, then it can be stated that the difficulty that children have in comparing quantities is

something very broad, more general, and more fundamental than number. The presentation of continuous material is unique to this investigation.

8.2 Summary and Integration of Experimental Findings

8.2.1 The Equalize and Compare Difference

It was hypothesized that Compare word problems would be much harder than Equalize word problems. Though, as previously mentioned, these two types of word problem involve the same mathematical operation, it was hypothesized that the same operation would be more difficult in the type of wording involved for a Compare word problem than in the type of wording involved for an Equalize word problem. This was explained by The Hypothesis which Combines Structural and Linguistic Factors (see Section 1.7 and 1.7.3).

In this investigation, the findings of Experiments 1, 2, 3, 4, and 5 support this Compare and Equalize difference. Children found Compare problems more difficult to solve than Equalize ones. However, this finding was not evident in Experiment 6. Experiment 6, however, was different than the previous experiments. The displays did not have the comparative pair represented in the choice stimuli. Children's performance on the Compare questions was improved by not having the comparative pair represented in the choice stimuli. This finding supports the Silver and Thompson Hypothesis, as well as the DasGupta and Bryant Representational Theory (see Section 1.7.1.3), as children were not forced back to an original value of one of the comparative stimuli when solving Compare-type questions.

8.2.2 The Non-Number-Specific Difference

This investigation was interested in finding out whether the children's patterns of responses would be confined to a number-specific context or would apply to quantity in general in order to find out how fundamental are children's difficulties with quantitative comparisons. Such an empirical-type question had never been asked before. Using both discontinuous and continuous quantities determined whether children made use of non-number-specific factors when solving word problems requiring comparisons of two unequal quantities.

Previous research by Briars and Larkin (1984) did not address the issue of whether children's responses to word problems was due to number or to quantity in general. Like, Riley (1981), Carpenter, Hiebert, and Moser (1981), DeCorte and Verschaffel (1987a), Riley, Greeno, and Heller (1983), Hudson (1983), and Silver and Thompson (1984), they were only interested in number. However, Briars and Larkin did suggest that word problem differences involved children's basic understanding of numerical concepts and number-specific comparisons.

In this investigation, the findings of Experiment 1 indicated that children made significantly more correct responses when presented with discontinuous material than with continuous material. However, the Equalize and Compare difference was evident in both these types of material. This indicated that the Equalize and Compare difference was not number-specific, but non-number-specific.

The results of Experiments 2 and 4, however, demonstrated a Question Type and Material interaction. This interaction indicated that children perform significantly better on Equalize questions than Compare when discontinuous material is used. Discontinuous material helps children's performance in Equalize questions, but not in Compare questions. These results disagreed with those of Experiment 1 where the Equalize and

Compare difference was significant with both discontinuous and continuous material. The Equalize and Compare difference is only significant with discontinuous material and not with continuous material in Experiments 2 and 4.

Experiments 3, 5, and 6 did not look at this material variable.

Hence, this investigation found that the difference between Equalize and Compare word problems was also evident using continuous quantities (Experiment 1). This indicated that children's difficulties in comparing quantities are non-number-specific. Such a finding had never before been reported.

It was also established that children found mathematical problems of an Equalize-type easier to solve when they were presented in a discontinuous fashion as opposed to a continuous fashion (Experiments 2 and 4). However, children found Compare-type problems difficult to solve when they were presented in both discontinuous and continuous fashion. Therefore, the difficulty children have with Compare problems applies to both discontinuous and continuous material. This indicated that children's difficulties in comparing quantities are number-specific, but their difficulty with Compare word problems is non-number-specific. Such a finding contradicts previous research by Briars and Larkin (1984) [see Section 1.7.1.2].

Hence, the Equalize and Compare difference is clearly evident in the experiments of this investigation. However, it seems that children represent the problems in two different ways, as Equalize problems seem to be number-specific and Compare problems seem to be quantity-specific. Such type results can be seen as taking place within the cognitive science or information processing framework. Before examining the results of this investigation in light of this approach, valid explanations of the results must be considered.

8.2.3 Valid Explanations of the Results

There seem to be three kinds of explanations to the above-mentioned results. One is a structural explanation, another is a linguistic explanation, and a third explanation is one which combines structural and linguistic factors.

8.2.3.1 The Structural Explanation

This first hypothesis was based on research by Riley, Greeno, and Heller (1983), who attributed the word problems' difficulty of children to a misrepresentation of problem structure. They further argued that children's difficulty with Compare word problems was due to their "static" characteristic. Children's relative ease with Equalize word problems was due to their "action" characteristic. Children found the latter easier than the former, as they can understand relationships involving action better than those that are static. These relationships involving action require a transformation to occur within its quantities by increasing or decreasing one of the quantities to make it equal to the other.

Therefore, from this investigation's results, there is evidence supporting a structural explanation to account for the difference between Compare and Equalize problems. The Equalize and Compare difference continued to emerge despite all the structural manipulations implemented in the experiments of this investigation. It was thought that by implementing these structural manipulations, the Equalize and Compare difference would not occur. These manipulations involved either making the Compare problems easier (Experiments 3 and 4) or making the Equalize problems difficult (Experiment 5).

The Equalize and Compare difference did not occur in Experiment 6. The lack of the Equalize and Compare difference was not due to the structural manipulation. However, this result only enhances the possibility of attributing the Equalize and Compare difference to a structural explanation rather than negating it. The children found the Compare problems easier to solve when the comparative pair was not represented in the choice stimuli. The children were then not forced back to an original value of one of the comparative pair. It seems that children were able to solve Compare problems by using a one-step process, rather than a multi-step process. The children did not have to keep in mind a representation of the initial quantity and, hence, were able to solve the problem quite successfully.

8.2.3.1.1 Structural Manipulations

Previous experiments were either about representation or about language (see Chapter One). Their results either supported the representational explanation or the linguistic one.

From the structural side, this investigation measured such variable in various ways. The evidence will now be reviewed in order to see to what extent it fits in with the explanations previously provided.

8.2.3.1.1.1 Part-Whole Manipulations

In an attempt to establish how difficult Compare word problems were to solve, manipulations in part-whole relationships (see Experiment 3) were implemented. The lack of interaction between Question Type and the part-whole Material indicated that this manipulation did not have any effect on children's performance of Equalize and Compare word problems. Hence, a

structural hypothesis of children's difficulties with part-whole relationships is not supported.

8.2.3.1.1.2 Story-Telling Context Manipulations

Story-telling context (see Experiment 4) was a further attempt at investigating the effects that different types of material could have on question type. The lack of interaction between Question Type and the Story-telling context indicated that this manipulation did not have any effect on children's performance of Equalize and Compare word problems. Hence, a structural hypothesis can not be explained in terms of different contexts.

8.2.3.1.1.3 Spatial Imagery Manipulations

8.2.3.1.1.3.1 Display Manipulations

Manipulations of display type (see Experiment 5) were implemented in order to examine the relative ease that children have with Equalize word problems. The lack of interaction between Question Type and Display Type indicated that this manipulation did not have any effect on children's performance of Equalize and Compare word problems. Hence, a structural hypothesis can not be explained in terms of children's difficulties with display.

8.2.3.1.1.3.2 Different-Level Display Manipulations

Different-level displays were introduced in order to pursue the question of spatial imagery. Results indicated that children were not generally affected by manipulations of spatial imagery when solving Equalize and Compare word problems (see Experiment 6). However, some evidence in support of the structural hypothesis was obtained.

8.2.3.2 The Linguistic Explanation and Its Manipulations

Research by Hudson (1983) attributed children's difficulty with Compare word problems to a misrepresentation of the comparative construction. (See Section 1.7.2.) This attribution to a general linguistic factor revolved around children's difficulty with the comparative word "more".

It was predicted, from this linguistic hypothesis, that there should be no difference between Equalize and Compare problems because both of them use the difficult term "more". However, in this investigation, this is not what happened (see Experiments 1, 2, 3, 4, and 5). The Equalize and Compare difference occurred when both problems involved the word "more". Therefore, a simple linguistic hypothesis about comparative terms is not viable.

A second question arises from the comparative terms issue. What about the differences between the comparative term "more" and the comparative term "less"? Furthermore, to what extent do these comparative terms account for the relationship between Equalize and Compare word problems?

In this investigation, the findings of Experiments 2, 3, and 4 demonstrated an interaction between Question Type and Comparative Term. Children performed better when the comparative term "more" was used than when the comparative term "less" was used. However, this was true only for the Equalize-type question and not for the Compare-type question; hence, the interaction. This difference, also, was true only for the younger group of children, between the ages of 5 and 6. The difference ceased to be true for the older group of children, between the ages of 7 and 8.

Therefore, it was confirmed that linguistic comparative markers affect children's performance in mathematical tasks, but not as Hudson would

predict. Linguistic markers thus proved to be developmentally affirmed and of not much consequence at the child's later stage. This being the case, and contrary to the theories of Hudson (1983), the linguistic aspect would then not be considered a crucial component for the properties of word problems, particularly not for the Compare-type problems.

Experiments 1 and 5 did not look at the comparative term variable.

However, a Question Type and Comparative Term interaction was not found in Experiment 6. It is interesting to note that the lack of this interaction coincided with the lack of the Question Type main effect. This result provides further evidence supporting the fact that a linguistic explanation cannot be looked at independently from the structural explanation in order to account for the difference between Compare and Equalize problems. It seems that the linguistic explanation does not hold without the structural explanation.

Therefore, the notion of a simple linguistic explanation to account for the Equalize and Compare difference does not seem feasible on its own. A more sophisticated and complicated language hypothesis in terms of its interaction with the structural task seems to be more appropriate to account for the Equalize and Compare difference. It was then considered that the Compare and Equalize difference can only be explained by a combination of structural and linguistic factors.

8.2.3.3 The Explanation which Combines Structural and Linguistic Factors

Research by Moore and Frye (1986) attributed children's difficulties to the two senses of the comparative term "more" (see Section 1.7.3). The word "more", according to Moore and Frye, can be used in two ways: Use (1) is to describe a change in a quantity over time, and Use (2) is to describe two particular quantities. The wording of an Equalize problem stresses a change

within a quantity, whereas the wording of a Compare problem stresses a comparison between two distinct quantities.

As previously addressed (see Section 1.7), the reason why the understanding of the two meanings of the word "more" is relevant to Equalize and Compare problems is that the wording in Equalize problems emphasizes Use (1), comparing the same quantity in two successive states, while in the Compare problems it emphasizes Use (2), comparing two different quantities without a transformation to either.

Children seem to understand the comparative term "more" in its Use (1), when it refers to an addition to one quantity, better than in its Use (2), when it refers to the difference between two quantities.

This research by Moore and Frye (1986) served as a precursor for this author's research, which attributes the word problems' difference to a combination of structural and linguistic factors. In this investigation, the findings of all experiments amass to provide support for this author's hypothesis (see Section 1.7, 1.7.3, and 1.8). In Experiments 1, 2, 3, 4, and 5, the Equalize and Compare difference occurred across rigourous manipulation of particular variables. These manipulations were aimed at finding out why this Equalize and Compare difference occurred. Experiment 1 established the relative ease that children have with Equalize word problems versus the evident difficulty that they have with Compare word problems. Experiment 2 further established, via the integration of the comparative term variables, that this Equalize and Compare difference cannot be explained by a simple linguistic hypothesis and that indeed the issue is a complex one crucial to children's solving of addition and subtraction problems. Experiment 4 provided further evidence for this conclusion as it dealt with different story-telling contexts via different comparative terms, establishing that the Equalize and Compare difference cannot be explained in terms of specific contexts. Experiment 3 established that the Equalize and Compare difference cannot

be explained via part-whole relationships by providing strong perceptual cues which divided the stimuli into parts. Experiment 5 established that the Equalize and Compare difference cannot be explained in terms of spatial imagery and perceptual factors by providing same-type and different-type material displays.

The Equalize and Compare difference was not evident in Experiment 6, which persisted on the manipulation of spatial imagery. It was particularly the negative results of this final experiment, which provided evidence for the hypothesis which combines structural and linguistic factors. Children's performance on Compare questions was improved by not having the comparative pair represented in the choice stimuli. Consequently, children's performance on the comparative term "less" was also improved by not having the comparative pair represented in the choice stimuli. It can thus, be concluded that the Equalize and Compare difference is determined by the "more" and "less" difference and vice-versa; the structural explanation is determined by the linguistic explanation and vice-versa. The amalgamation of the structural and linguistic factors as the explanation to children's difficulty when solving Compare word problems and their relative ease when solving Equalize word problems is inevitable in light of these results.

Furthermore, the choice of strategies provide additional support for this author's hypothesis. Children were not able to use the Equivalence Strategy to arrive at the solution. The Equivalence Strategy, popularly used in Compare problems supports the structural explanation that Compare problems embody a static characteristic. The Equivalence Strategy is, in itself, static by nature as all it entails is choosing, of the choice stimuli, the equivalent in height of the larger standard or choosing the equivalent in height of the smaller standard to arrive at a solution. Such strategy does not require any quantity transformation involving addition or subtraction to make one quantity equal to the other, as required in action relationships. Such a

strategy supports the structural explanation that relationship and representation are key issues in explaining the results obtained.

The choice of strategies also provides support for the linguistic explanation. The Equivalence Strategy provides evidence for children's understanding of the comparative term "more" in its Use (2) and not in its Use (1). By definition, the Equivalence Strategy is not representative of understanding an addition over and above the amount already present, which is the Use (1) of the word "more".

Hence, the lack of a Question main effect and a Question Type and Comparative Term interaction in Experiment 6 may be due to the fact that children were not able to use the Equivalence Strategy. This provides further evidence that the only real explanation accounting for the Compare and Equalize difference is the hypothesis which combines structural and linguistic factors.

8.2.3.3.1 Analysis of Strategies

An even better way of finding support for the hypothesis which combines structural and linguistic factors is to analyze the different types of strategy that children use when solving Equalize and Compare problems. The data on strategies emphasize the qualitative distinction between Compare and Equalize problems (see Experiments 2, 3, and 4).

It can be concluded that the difference between Equalize and Compare problems varies across both discontinuous and continuous material. It was confirmed, in this investigation, that children use the known additive and subtractive strategies in the Equalize problems with discontinuous material. However, it was found (see Experiments 2, 3, and 4) that when this type of problem was given with continuous material, children invented their own strategies, homogeneous in nature to the ones they used

for discontinuous material. Correct strategies were used more frequently for the Equalize question than for the Compare question. Whatever strategies children possess and use for the Equalize-type problem, they were not capable of using them to solve the Compare-type problem.

Children who used cognitive representation strategies (Counting Strategy or Measuring Strategy [see Experiments 2, 3, and 4]) usually arrived at the correct solution of the word problem. Cognitive representation strategies require an integration of two types of understanding: [(1) An understanding of absolute amounts and (2) An understanding of a relational amount] prior to working out the problem. Children who instead used perceptual strategies (Equivalence Strategy [see Experiments 2, 3, and 4]) usually arrived at the incorrect solution.

Equivalence strategies do not require an integration of the above-mentioned types of understanding in order to reach the solution of a word problem. They do not require a cognitive representation of the problem. Instead, they form part of a perceptual representation of the problem, choosing the choice stimuli which is the equivalent in height of the larger or smaller standard. This strategy entails choosing an absolute amount which is already represented in the problem.

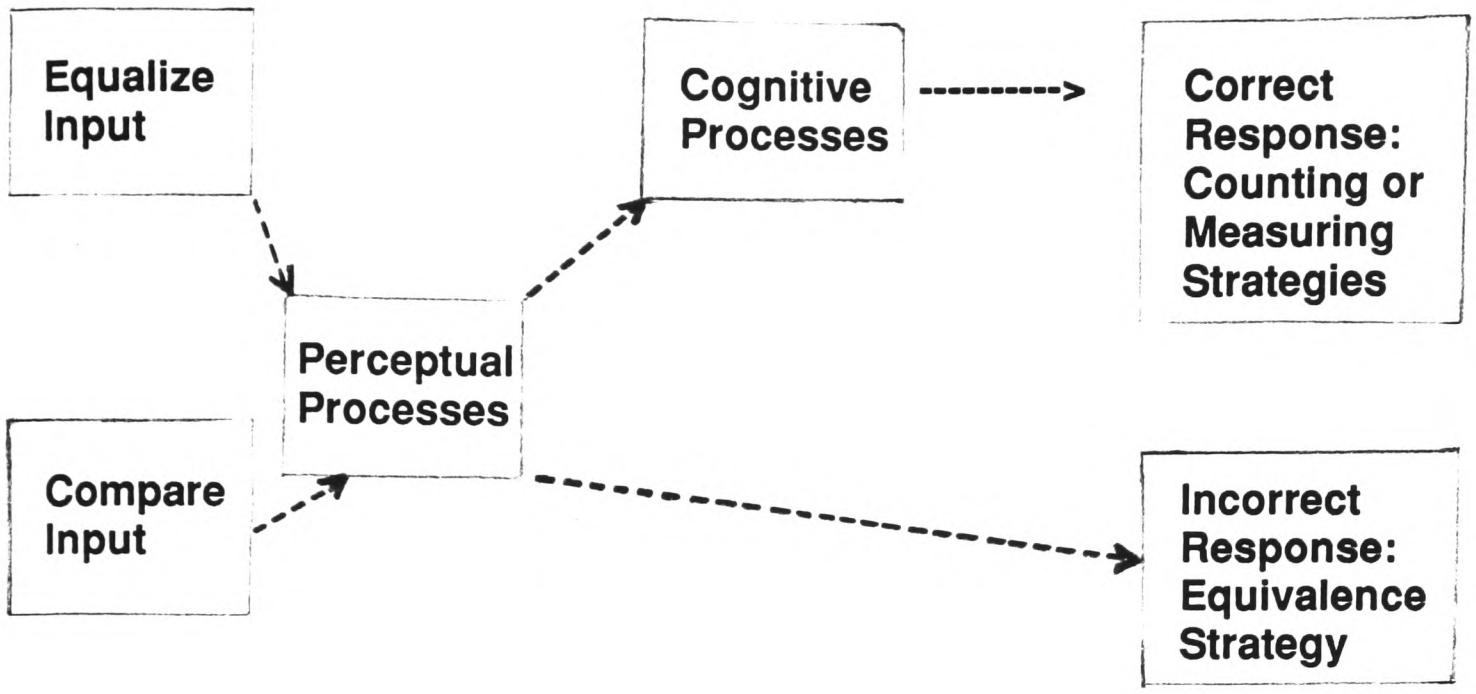


Figure 7. Visual representation of using Cognitive representation strategies versus Equivalence strategies to arrive at either a correct or incorrect solution to Equalize and Compare word problems.

The above graphical figure illustrates that in order to arrive at a correct solution in word problem solving, children should use cognitive representation strategies. In many instances, children make use of such strategies for the Equalize-type problems and arrive at the correct solution. However, in most instances, children do not use cognitive strategies for the Compare-type problems. They use Equivalence strategies and arrive at an incorrect solution.

It is suggested that, as the difference between Equalize and Compare word problems exists not only in the way these two type problems are represented, but also in the choice of strategies children use to solve each type of problem, an information-processing approach would provide a broader description of how the information is processed and represented in the child's mind. Before investigating how the hypothesis which combines structural and linguistic factors, as the explanation to the Equalize and Compare difference, fits into a more general social cognition type research,

other related explanations to the Equalize and Compare difference must be examined.

8.2.3.4 Other Related Explanations

8.2.3.4.1 The Silver and Thompson Hypothesis

Silver and Thompson's (1984) hypothesis generally dealt with the fact that different word problems require a different number of steps to arrive at a solution (see Section 1.7.1.3). The fact that Compare word problems are more difficult is, according to this hypothesis, due to the number of steps required in arriving at the solution of the above-mentioned problem. This hypothesis may be the same as the Riley, Greeno, and Heller's (1983) structural hypothesis, but couched in different terms. Silver and Thompson did not apply their hypothesis directly to Equalize and Compare word problems.

However, their hypothesis can be lifted from a word problem setting to apply to an even more general issue, such as that one of quantity. An Equalize problem involves an action relationship where there seems to be a 1-step process to arrive at the solution (from an absolute value to a relational value). In the experiments in this investigation, by just looking at the two endpoints the child can cancel out the two quantities against each other to arrive at the difference.

A Compare problem, on the other hand, involves a static relationship which seems to require a 2-step process to arrive at the solution (from an absolute value to a relational value and back to an absolute value). The child has to look at the end point of Quantity A and know how much there is in Quantity A. The child then has to proceed to look at the end point of Quantity B and know how much there is in Quantity B. Then, the child has to infer how

much Quantity A was initially (on the basis of Quantity B) in order to decide a change and arrive at a solution.

An end-anchor known value is present when the given value at one end of the array becomes crucial in the Compare-type question because the answer is based on an absolute value, and not on a relational one as it is in the Equalize-type question.

8.2.3.4.2 The DasGupta and Bryant Hypothesis

Two-step series problems, such as Compare problems, have generated considerable debate concerning the nature of representation. In the present investigation, it may be the case that for an Equalize-type problem the child only used the endpoints of Quantities A and B to arrive at the solution, but for a Compare-type problem, the child had to refer back to an initial state of Quantity A in order to see what the difference between the two quantities actually was.

The parallel drawn with recent work on causal reasoning in children by DasGupta and Bryant (1989) [see Section 1.7.1.3] established the possibility that 5-to-6-year-olds are unable to refer back to an original starting point which is not actually represented in Compare word problems. It seems that children do not have to keep in mind representations for Equalize word problems, but need to have the ability to infer representations of an initial quantity and keep them in mind for a Compare word problem.

8.2.3.4.3 The Pirie and Kieren Hypothesis

A parallel can also be drawn with recent work on mathematical understanding by Pirie and Kieren (1989; 1990; 1991). Pirie and Kieren developed a hypothesis about mathematical understanding. Their hypothesis

is not about word problems in general and does not deal specifically with the Equalize and Compare difference, but nevertheless, can be applied to this difference.

The Pirie and Kieren Model is concerned with the growth of mathematical understanding. This understanding is seen as a whole dynamic process with non-linear knowledge categories. The model has a recursive structure which features the growth of understanding as a "folding back" process.

Hence, there are eight embedded layers to this model. Each layer represents a knowledge category whereby knowing grows recursively using representations made on previous knowing. These levels do not constitute understanding independently, but are named parts of the whole phenomenon of the growth of understanding.

Pirie and Kieren's Model

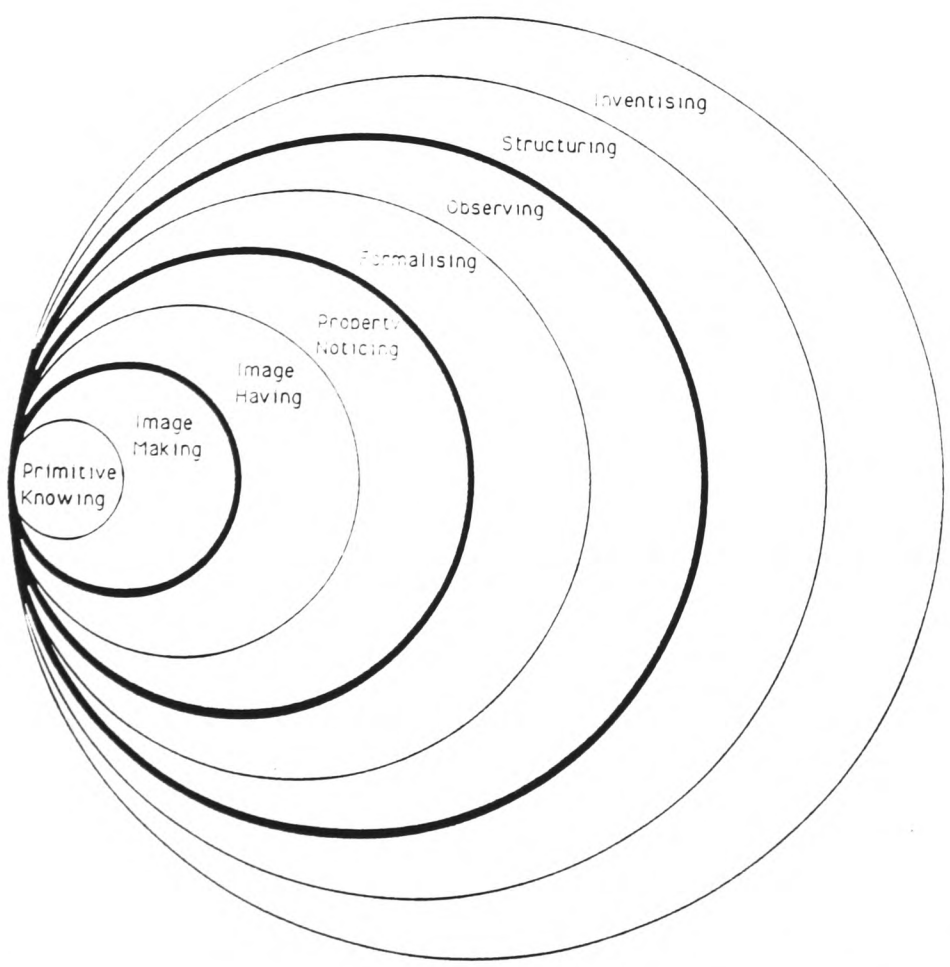


Figure 8. Example of Pirie and Kieren's (1991) growth of mathematical understanding model.

The first of the eight embedded layers is named Primitive Knowing. This first layer refers to everything that is in the mind to start a new concept. The second of the eight embedded layers is Image Making. This second layer refers to a mental image having been formed by a physical doing or activity. The third of the eight embedded layers is Image Having. This third layer refers to a stronger image of the concept without having to do anything to form it. Singular actions are now represented by mental images. The fourth of the eight embedded layers is Property Noticing. This fourth layer refers to mental images as they are examined by mental properties. Images are examined to see what can be said about their specific and relevant properties. The fifth of the eight embedded layers is Formalizing. This fifth layer refers to the knowledge of class-like mental objects; that is, abstracting common qualities. The sixth of the eight embedded layers is Observing. At this layer, one can step back from formalization and organize these formal thought structures. The seventh of the eight embedded layers is Structuring. It is at this level where number is driven logically. One becomes aware of associations, as well as their sequence and interdependence within one's observations. The last layer is Inventizing. This layer refers to having a complete structured understanding, including a knowledge of what happens in case of change.

Pirie and Kieren have used this two-dimensional representational model to explain how individuals understand mathematics. Their model is based on their view that mathematics is a sequential topic which is taught in a sequential fashion. However, no one learns it in that manner. They claim that understanding is achieved only by making connections through the layers. These connections are made by means of a process and completed with a form. Having a process at a layer is a precursor to building a form for that layer. Hence, their recursive model, permits a "folding back" of layer(s) in order to make the connections.

Pirie and Kieren have hencefar only used their model to explain a person's working within an observed conceptual area. They have not expanded their theory to apply to a group of people's understanding of a certain topic, nor to apply to how particular topics are understood. Their model is domain-specific to the individual and not domain-general. The possibility exists, however, that their model can be mapped to the Equalize and Compare difference in a general way.

The possibility is established that within their model of the growth of mathematical understanding children performing an Equalize-type problem do not need to "fold back" in order to arrive at the correct solution. For example, in the case of the Equalize problem in the experiments in this investigation, a child looks at the two endpoints (Primitive Knowing), then the child can cancel out the two quantities against each other (Image Making) to arrive at the difference (Image Having).

Children performing a Compare-type problem "fold back" twice in order to arrive at the correct solution. For example, in the case of the Compare problem in the experiments in this investigation, a child has to look at the end point of Quantity A (Primitive Knowing) and know how much there is in Quantity A (Image Making). The child then has to proceed to look at the end point of Quantity B ("fold back" to Primitive Knowing) and know how much there is in Quantity B (Image Making). Then, the child has to infer how much Quantity A was initially, on the basis of Quantity B, ("fold back" again to Primitive Knowing) in order to decide a change (Image Making) and arrive at a solution (Image Having).

The process of understanding Equalize and Compare problems is different within the Pirie and Kieren model. For both Equalize and Compare problems, Image Making involves knowing the absolute values. A relational value is also achieved at this layer as it requires an activity in order to get the idea of the concept. A relational value is a comparison between two

quantities involving multiple actions, but it is still an activity. Children seem to arrive at the same level of understanding, Image Having, for both these type problems, but the process by which they arrive at this level is different. Image Having is the third out of the eight embedded layers for mathematical understanding. This is the level at which singular actions can be presented by mental images and at which a mental object can now be used in mathematical knowing. At this level, a child might know, for example, that "the response set" shows the amount that comes from their use of cognitive representational strategies.

There seems to be a parallel with the Silver and Thompson (1984) hypothesis with regards to the number of steps to arrive at a solution. These may be considered equivalent to the amount of "folding back" before arriving at the correct solution in the Pirie and Kieren hypothesis. The Equalize problem seems to involve a 1-step process or no "folding back" to arrive at a solution, whereas the Compare problem seems to involve a 2-step process or "folding back" twice to arrive at a solution. Hence, once again, performance on Equalize word problems is at a striking advantage over performance on Compare word problems based on the processing demands for mathematical understanding in the Pirie and Kieren model. It is easier to form an image for the Equalize word problems than it is for the Compare word problems.

8.2.3.4.4 The Absolute versus Relative Amount Hypothesis

Bryant (1974) proposed that children understand relative codes, but have difficulty with absolute codes. Bryant suggests that this is the case along any continuum.

This claim can be applied to the results of this thesis. Children may be able to understand word problems which require their solving through relative

codes, but will have difficulty with word problems which require their solving through absolute codes.

In accordance with the hypothesis of this author and other related hypotheses, such as those of Silver and Thompson (1984), DasGupta and Bryant (1989), and Pirie and Kieren (1989; 1990; 1991), it is proposed that arriving at the solution of an Equalize word problem involves solving the problem using relative codes. It is also proposed that arriving at the solution of a Compare word problem involves solving the problem by depending on the absolute amounts.

In the experiments of this thesis, it is suggested that a child solves an Equalize problem, for example, by only looking at the two endpoints of Quantity A and B. The child arrives at the solution by depending primarily on relative codes.

In the experiments of this thesis, it is suggested that a child solves a Compare problem, for example, by looking at the endpoint of Quantity A and knowing how much there is in Quantity A (registering absolute amount [1]); the child then has to look at the endpoint of Quantity B and know how much there is in Quantity B (registering absolute amount [2]); the child then has to refer back to the initial Quantity A in order to decide a change and arrive at a solution (arriving at a relative amount based on registered absolute amounts).

Hence, it is not surprising that in the experiments of this thesis, where two age groups participated (see Experiments 1, 5, and 6) the older children performed much better than the younger children. In Experiment 1, young children found it particularly difficult to solve Compare word problems. Hence, there was a Question x Age interaction. They could not arrive at the solution to this problem because it required their use of absolute codes. However, they did solve Equalize word problems with some ease as these required their use of relative codes.

In Experiment 5, where Equalize questions were made harder by manipulating spatial imagery, young children generally performed worst than the older children. Hence, there was a main effect of Age, but no interaction between Question x Age. In Experiment 6, the absolute amounts of the comparative pair were not represented in the choice stimuli. Children were not able to use an absolute code in order to arrive at the correct solution of the word problem. Compare problems were easier to solve when the absolute amounts were not represented in the choice stimuli. The children were not forced to use absolute codes in order to solve the problem. The lack of Question main effect and lack of Question x Age interaction indicated that this improvement in Compare problems was true for both age groups.

Hence, the successful solving of word problems may depend on a developmental change from relative to absolute. This developmental change seems to be inherent in the structural and linguistic properties of the two types of word problems.

8.2.3.4.4.1 The Role of The Equivalence Strategy in The Absolute versus Relative Amount Hypothesis

The Equivalence Strategy plays a significant part in Bryant's (1974) hypothesis. Bryant claimed that children are able to recognize "broad" relationships. These "broad" relationships entail recognizing the "larger" or the "smaller" in a given relationship.

The Equivalence Strategy does not represent the child's understanding of absolute amount, but rather their knowledge of "more" and "less", "bigger" and "smaller", "taller" and "shorter". This knowledge is necessary before an understanding of a relational amount is achieved and before an understanding of absolute amount is achieved.

Hence, the Equivalence Strategy demonstrates some sign of an understanding of relative codes and a transition ground before the understanding of relational amounts. Bryant emphasized the limitations of relative codes. The limitation of the Equivalence Strategy is quite obvious. Children seem to use it only when they find a problem too difficult to process and henceforth, arrive at the incorrect solution.

It also seems that the initial understanding of the Equivalence Strategy can be related to perceptual understanding. As previously noted (see Section 8.2.3.3.1), equivalence strategies form part of a perceptual representation of the problem and do not require a cognitive representation of the problem.

8.3 Recommendations

8.3.1 Implications for Teaching

The results of Experiment 1 indicated that the Equalize and Compare difference occurs as much with discontinuous as with continuous quantity. This phenomenon may have implications for teaching and about the way children should learn to compare differences in number. Since the difficulty children have in comparing quantities is not specific to number, but more fundamentally, to quantity in general, as it applies to both discontinuous and continuous quantities, children may be encouraged to sort out quantitative problems separately from number. This should be done prior to the teaching of addition and subtraction.

Experiment 1 made use of concrete material. It is then suggested that teachers design examples of problems involving quantitative tasks, such as those involving measurement tasks, with concrete material, in order to help the child sort out quantitative differences before numerical differences. Only then will a child be better prepared to handle numerical comparisons.

Furthermore, the strong result regarding the order of the questions in Experiments 3 and 4 may have implications for teaching. Success at Compare-type questions was affected by whether or not they were presented after the Equalize-type questions. This was not the case with Equalize-type questions. They were not affected by whether or not they were presented after the Compare-type questions. There are two possible ways of looking at the implications of this order effect.

(1) Implications for Drill Teaching

It is suggested that children be drilled on Equalize-type word problems first. They must be able to successfully establish an understanding of this type problem before they are presented with Compare-type word problems. This has strong educational implications as drilling children on Equalize-type questions may then trigger successful performance on Compare-type questions.

(2) Implications for Sequential Teaching

It was suggested that there may exist a developmental sequence in the acquisition of the above-mentioned problems (see Section 8.2.3.4.4). The order effect suggests that there is a developmental trend prior to arriving at a successful solution. This developmental trend is seen in (a) the understanding of the problems and (b) the use of strategies in each of the problems.

The developmental trend (a) is seen by the fact that Equalize word problems are understood before Compare word problems. The developmental trend (b) is seen by the fact that children make use of the Equivalence Strategy before they make use of the Counting and Measuring strategies. The Equivalence Strategy, which does not yield a correct solution, seems to develop before the Counting and Measuring strategies, which do yield a correct solution (see Section 3.4).

Hence, it does seem, from the above-mentioned, that one can generalize from children's performance on Equalize word problems to their eventual performance on Compare word problems. Their understanding of Equalize word problems and their use of strategies in such type problems will lead to their understanding of Compare word problems and their use of strategies in these type problems. The teacher should make further use of this evidence in considering the teaching of Equalize word problems prior to the teaching of Compare word problems.

8.3.2 Directions for Future Research

Future experiments could be designed to further test the hypothesis which combines structural and linguistic factors. It is suggested that this be done by designing similar experiments to those described in this thesis, by varying one aspect of the continuous displays: the position of the items. The bars, lines, and men described in the experiments were all in vertical positions. It is possible that rotating the position to a horizontal one would alter the results. (See Figure 9.) Another variation could be done only in the choice stimuli. One could alter the positions of the items in the choice stimuli randomly. Some could be in a vertical position, others in a horizontal position. (See Figure 10.)

Experiment 5 was designed to make the Equalize problem difficult by providing same-type and different-type display manipulations. It was thought that by implementing this spatial imagery manipulation, the Equalize and Compare difference would not occur. The results of this experiment, nevertheless, yielded the Equalize and Compare difference, indicating that this manipulation did not work. It is thought that these further manipulations of spatial imagery may provide interesting results. Once again, the aim of these manipulations would be to make the Equalize questions difficult. It is

hypothesized that the Equalize and Compare difference would not occur. This would provide further evidence for this author's hypothesis.

The lack of a significant display effect suggests that children were not able to use an aid to help spatial imagery moves to arrive at the correct solution in Experiment 6. There was also a surprising absence of a question effect which showed no significant difference between Equalize and Compare problems. However, it was speculated that the lack of Equalize and Compare difference in this experiment may have been due to the fact that the choice stimuli contained no absolute value as the comparative pair. It is thought that a further manipulation of this particular representation may provide interesting results. In fact, an experiment is needed where in one condition, the comparative pair is represented in the choice stimuli, and in another condition, it is not. In this way, a comparison across conditions can be made in order to check out the effect of the Equivalence Strategy. If the Equalize and Compare difference does not occur, this author's hypothesis would be further supported.

Further research can also be done via longitudinal studies. Since the ease children have with Equalize problems is number-specific and the difficulty children have with Compare problems is non-number-specific, it would be interesting to test children, doing the same tasks in this investigation, at ages three and four for the Equalize questions and at ages eight and nine for the Compare questions. This further research would be done in order to see at what developmental stage are Equalize questions quantity-specific and, respectively, at what developmental stage are Compare questions number-specific. One would be able to find out whether this Equalize and Compare difference is more general to a developmental-cognitive science on information-processing. This follow-up study would inevitably recount a child's history and development of how s/he has been

comparing unequal quantities. There would be much to gain from this type of longitudinal research.

Other future experiments could involve the use of intervention methods. As this investigation has proposed, quantitative comparisons should precede numerical comparisons and Equalize word problems should precede Compare word problems (see Section 8.2.1), it would be very interesting to design experiments to test the overall efficiency and success of such implementation. Again, there would be much to gain from this type of intervention research.

Finally, it is proposed that these same experiments be carried out using all four different types of word problems. It would be very interesting to find out whether there are any additional differences that would come up with regard to Change and Combine word problems and whether these additional differences would be as great with discontinuous as with continuous material. This type of research would not only enable us to understand how children compare quantities and find out the difference between them in all possible established ways, but also to see whether these different-type word problems trigger different information-processing mechanisms. A complete proposal on how to teach word problem solving to children 5--to-8-years-old, could be drawn based on this research.

8.3.3 Word Problems in an Information-Processing Capacity

It is now essential to look at the implications of this investigation's results in a broader, more general theoretical framework, as is the information-processing approach. Information-processing regards the mind as a complex cognitive system. Metaphorically speaking, information-processing regards the mind as a computer, whereby it manipulates or processes information obtained from the environment or information already

stored within. Flavell (1985) described some processes of information-processing such as encoding, recoding, and decoding; comparing and combining with other information; storing in or retrieving from memory; and bringing into or out of focal attention or conscious awareness. The information-processing approach provides a theory of cognitive development that: (1) considers the resources available in the cognitive system, (2) considers the way the information is selected, and (3) considers the way the information is manipulated.

Theories of information-processing prospered after the decline of Piaget's influence on theories of cognitive development. Piaget regarded cognition as a system working according to general principles. Information-processing served to give more precise and exact specification of what information is relevant to the child when performing a particular task, as well as how the child manipulates and stores this information. Hence, the information-processing approach is not an alternative to Piaget's theory of cognitive development (McShane,1991). There is not one grand theory of information-processing, but rather numerous information-processing theories of cognitive development, depending on the task at hand. These information-processing theories involve computer-simulation models in order to account for specific processes underlying task performance. Such models emphasize precisely how information is transmitted and manipulated through the cognitive system and specify the precise steps that are involved in the processing of applying rules to information.

Hence, the information-processing approach follows the basic functions of any cognitive system: receiving, processing, storing, and retrieving information. It is the information in question which is specified by the task at hand. A child will form representations of the available information, which is known as the input. There are two basic types of research within the information-processing framework. One type of research

is concerned with the information-processing system itself which is what actually develops in the system. Research of this type has focused on basic perceptual processes or memory processes. These are the control processes that manipulate the information. The second type of research is concerned with the performance on particular tasks. Task performance in the information-processing approach is considered to be the application of a certain set of rules to a particular task. A rule is a procedure or strategy that the child uses to perform the task. These rules explain children's performance on a given task, that is, their errors and correct choices.

With regard to reasoning about quantity and number, Piaget and Inhelder (1971) saw mathematics, logic, and reasoning as interrelated. Before numerical calculations are carried out, children must first consider the logical structure and reasoning behind the problem. Once the child builds logical relationships, then numerical concepts can be introduced.

The information-processing approach contrasts with Piaget's theory of cognitive development in various ways. Whereas Piaget assumed mathematics, logic, and reasoning to be interrelated, thus claiming that the same underlying cognitive structures are responsible for the understanding of the world, the information-processing approach claims that the cognitive system is made up of autonomous structures which are not interrelated and are responsible for the understanding of different aspects of the world. These different aspects of the world constitute domains. A domain is a collection of tasks that share a common representational system and a common set of procedures. The cognitive system then supports a wide range of representational structures. Some of these representational structures require dedicated processes for their manipulations. It is this latter view that is able to explain why children demonstrate better performance in one area than in another. The information-processing approach explains task performance within a domain of knowledge as requiring independent and

specific procedures, as well as independent and specific systems of representation. Different types of word problem seem to require the development of specialized representations and procedures to create a cognitive domain.

As previously reviewed, the representational systems that young children at first use to solve Equalize and Compare word problems are indeed quite different, as are the procedures involved. The hypothesis that combines structural and linguistic factors, which explains the Equalize and Compare difference, emphasizes these different representational systems and procedures. Hence, it seems that Equalize and Compare problems may exist in two different domains as far as young children are concerned. These two different domains have different representational systems, each with their own dedicated processes for manipulating these representations.

The information-processing approach may be a more appropriate way of explaining the results obtained in this investigation in a broader theoretical context. It is quite possible that this approach will explain why children find and adopt completely different strategies for Equalize and Compare problems.

DIRECTIONS FOR FUTURE RESEARCH: FIGURES

Directions for Future Research

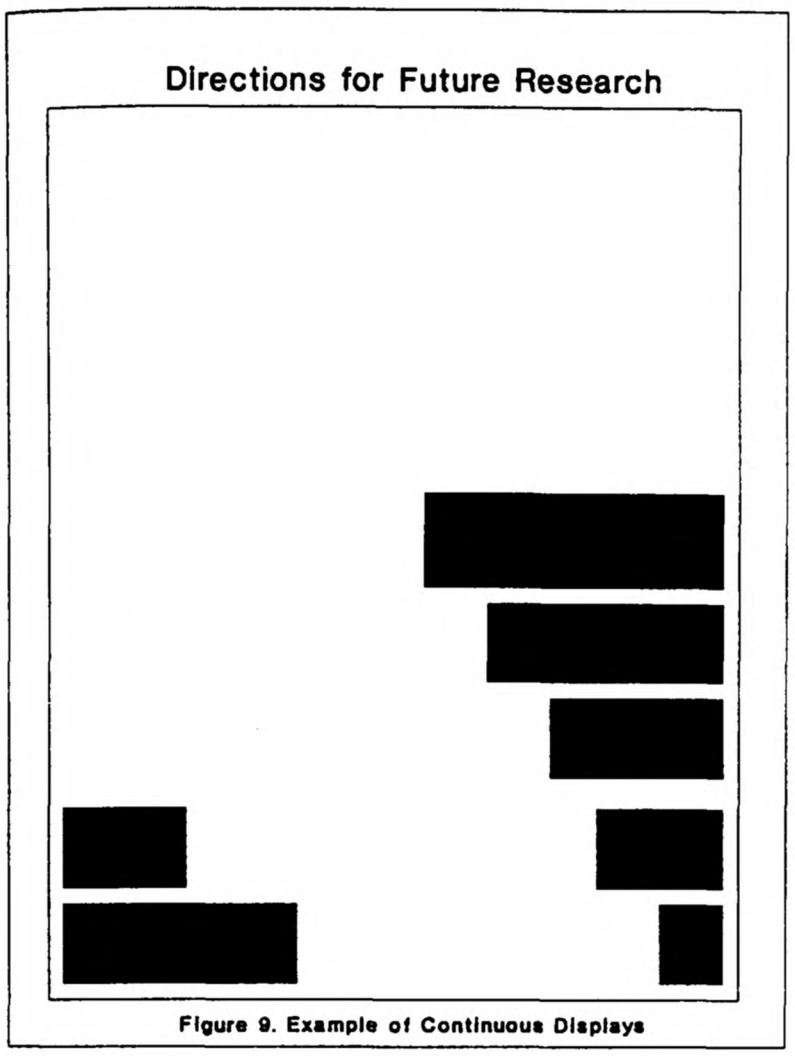


Figure 9. Example of Continuous Displays

Directions for Future Research

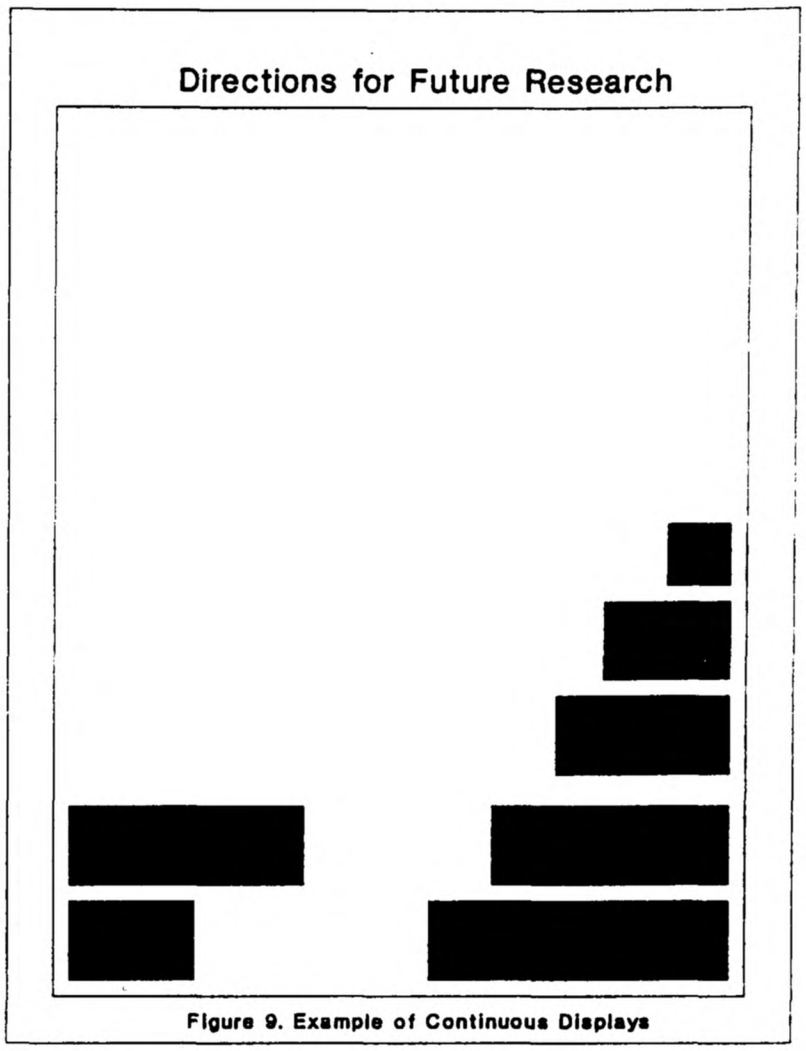


Figure 9. Example of Continuous Displays

Directions for Future Research

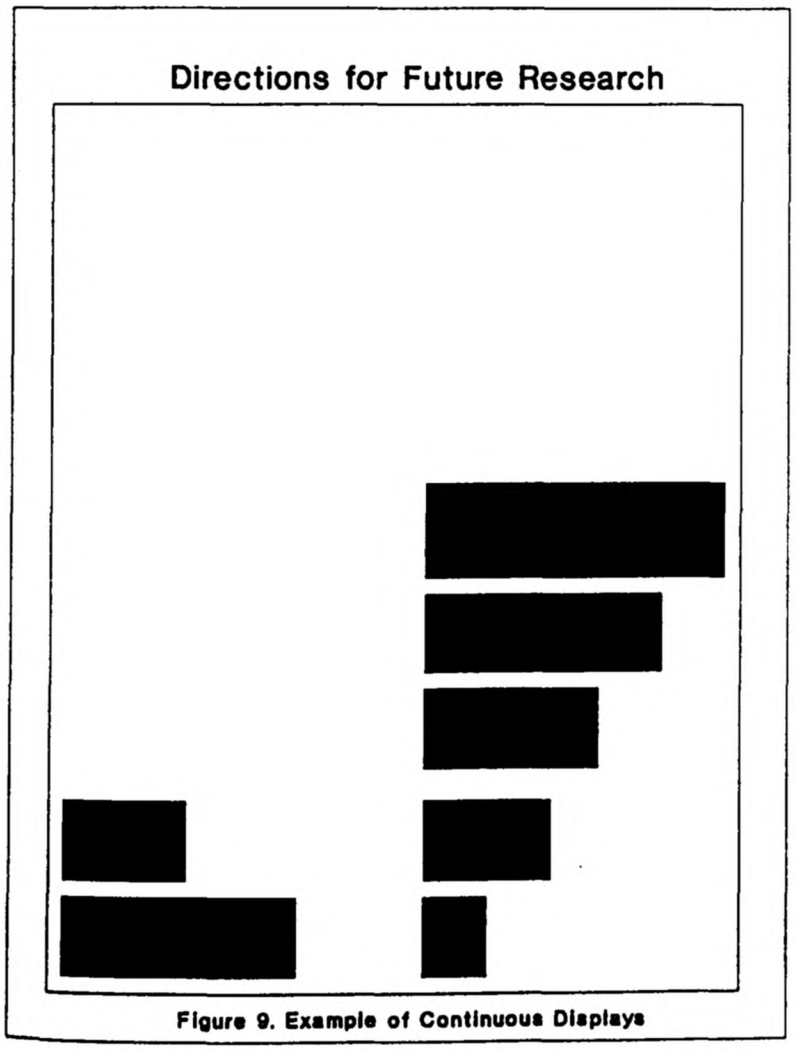


Figure 9. Example of Continuous Displays

Directions for Future Research

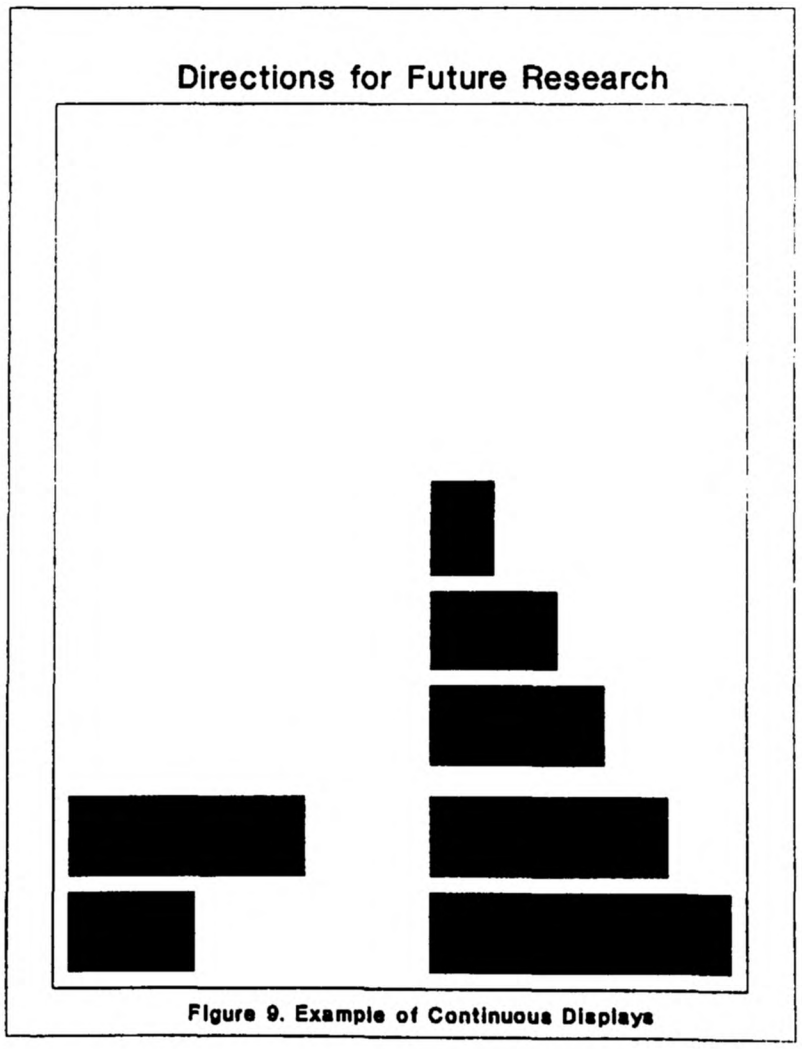
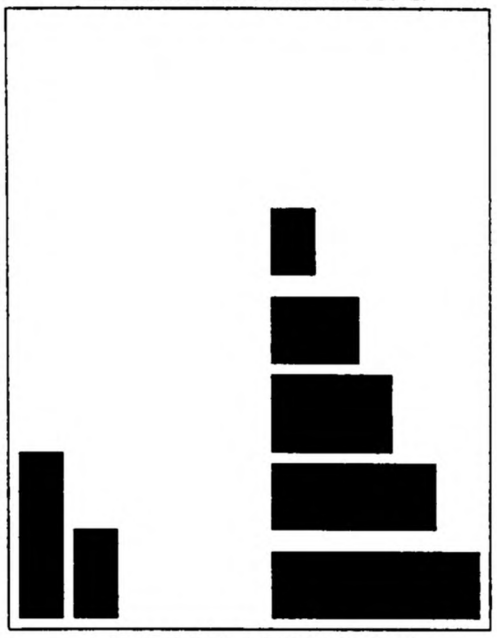
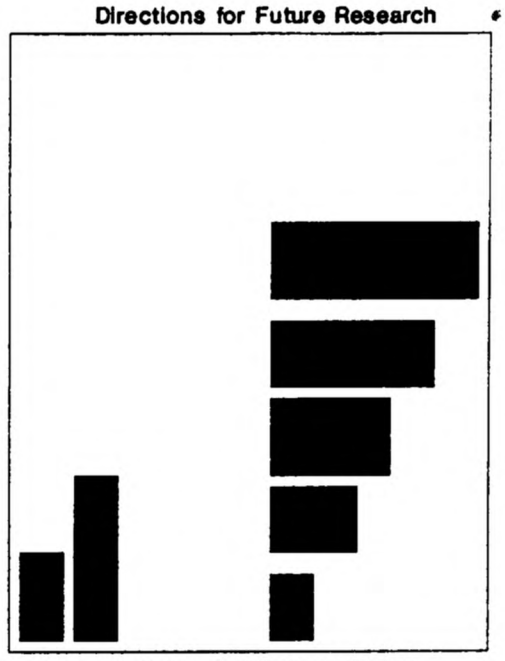


Figure 9. Example of Continuous Displays



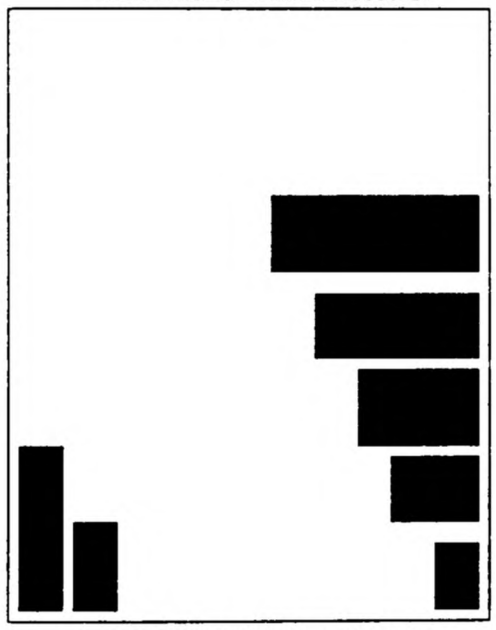
Directions for Future Research

Fig 10. Example of Continuous Displays



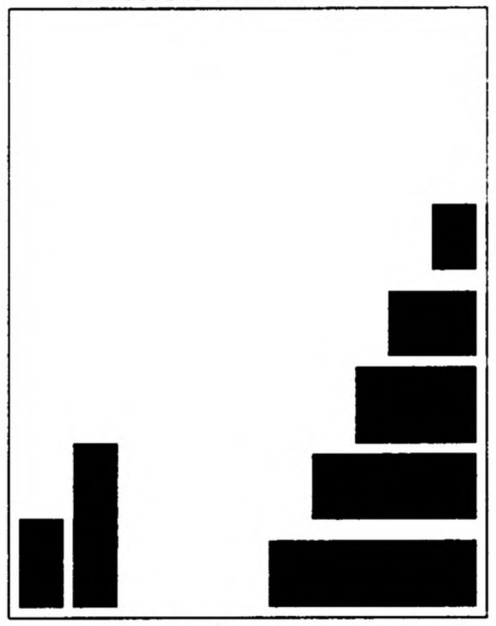
Directions for Future Research

Fig 10. Example of Continuous Displays



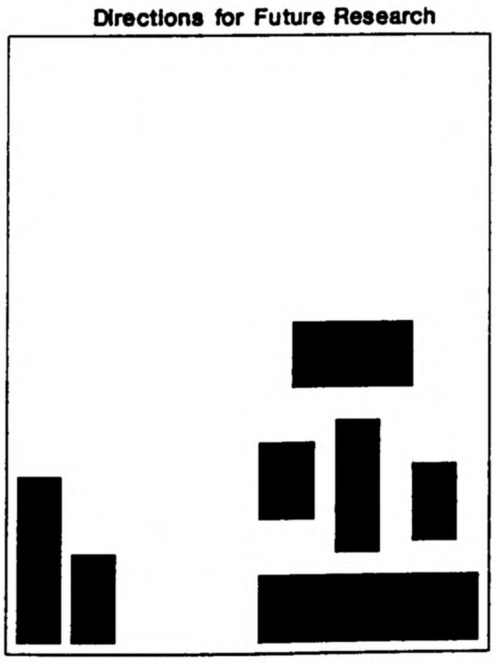
Directions for Future Research

Fig 10. Example of Continuous Displays



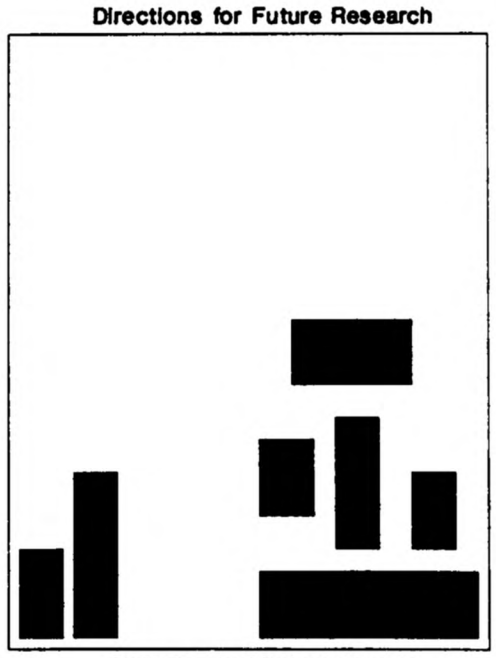
Directions for Future Research

Fig 10. Example of Continuous Displays



Directions for Future Research

Fig 10. Example of Continuous Displays



Directions for Future Research

Fig 10. Example of Continuous Displays

REFERENCES

Ballew, H., & Cunningham, J.W. (1982). Diagnosing Strengths and Weaknesses of Sixth-Grade Students in Solving Word Problems. *Journal for Research in Mathematics Education*, 13(3), 202-210.

Baroody, A.J. (1983). The development of procedural knowledge: An alternative explanation for chronometric trends of mental arithmetic. *Developmental Review*, **3**, 225-230.

Baroody, A.J. (1984a). Children's Difficulties in Subtraction: Some Causes and Cures. *Arithmetic Teacher*, **32**(3), 14-19.

Baroody, A.J. (1984b). Children's Difficulties in Subtraction: Some Causes and Questions. *Journal for Research in Mathematics Education*, **15**(3), 203-213.

Baroody, A.J. (1984c). *The Development of Counting Strategies for Single-Digit Addition.* Departmental working paper, Graduate School of Education and Human Development, University of Rochester, Rochester, New York.

Baroody, A.J. (1984d). More precisely defining and measuring the order-irrelevance principle. *Journal of Experimental Child Psychology*, **38**, 33-41.

Baroody, A.J., & Ginsburg, H.P. (1983). The effects of instruction on children's concept of "equals". *Elementary School Journal*, 84, 199-212.

Baroody, A.J., & Ginsburg, H.P. (1986). The Relationship Between Initial Meaningful and Mechanical Knowledge of Arithmetic. In Hiebert, J. (Ed.), *Conceptual and Procedural Knowledge: The Case of Mathematics*. Hillsdale, New Jersey: Lawrence Erlbaum Associates.

Bloom, L. (1970). *Language development: form and function in emerging grammars*. Cambridge, Massachusetts: MIT Press.

Brainerd, C.J. (1973). Mathematical and behavioural foundations of number. *Journal of General Psychology*, **88**, 221-281.

Brandimonte, M.A., Hitch, G.J., & Bishop, D.V.M. (1992). Manipulation of Visual Mental Images in Children and Adults. *Journal of Experimental Child Psychology*, **53**, 300-312.

Briars, D., & Siegler, R.S. (1984). A featural analysis of preschoolers' counting knowledge. *Developmental Psychology*, 20, 607-618.

Briars, D.J., & Larkin, J.H. (1984). An Integrated Model of Skill in Solving Elementary Word Problems. *Cognition and Instruction*, 1(3), 245-297.

Bruning, J.L., & Kintz, B.L. (1977). *A Computational Handbook of Statistics*. Glenview, Illinois: Scott, Foresman, and Co.

Bryant, P.E. (1972). The understanding of invariance by very young children. *Canadian Journal of Psychology*, 26, 78-96.

Bryant, P.E. (1974). *Perception and Understanding in Young Children - An Experimental Approach*. London: Methuen & Co. Ltd.

Carpenter, T.P., & Moser, J.M. (1982). The development of addition and subtraction problem-solving skills. In Carpenter, T.P., Moser, J.M., & Romberg, T.A. (Eds.), *Addition and Subtraction: A Cognitive Perspective*. Hillsdale, New Jersey: Lawrence Erlbaum Associates.

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

Carpenter, T.P., Corbitt, M.K., Kepner, H.S., Lindquist, M.M., & Reys, R.E. (1980). Solving verbal problems: Results and implications from national assessment. *Arithmetic Teacher*, 28(1), 8-12.

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), 27-39.

Carpenter, T.P., Lindquist, M.M., Silver, E.A., & Matthews, W. (1984). Achievement in mathematics: Results from the national assessment. *Elementary School Journal*, 84(5), 485-495.

Cowan, R. (1987). Assessing children's understanding of one-to-one correspondence. *British Journal of Developmental Psychology*, 5(2), 149-153.

DasGupta, P., & Bryant, P.E. (1989). Young children's causal inferences. *Child Development*, 60, 1138-1146.

DeCorte, E., & Verschaffel, L. (1985). Beginning first graders' initial representation of arithmetic word problems. *Journal of Mathematical Behaviour*, 4, 3-21.

DeCorte, E., & Verschaffel, L. (1987a). Children's problem-solving skills and processes with respect to elementary arithmetic word problems. In DeCorte, E., Lodewijks, H., Parmentier, R., & Span, P. (Eds.), *Learning and instruction. European research in an international context*. Oxford/Leuven: Pergamon Press/ Leuven University Press.

DeCorte, E., & Verschaffel, L. (1987b). 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.

DeCorte, E., Verschaffel, L., & DeWin, L. (1985). Influence of Rewording Verbal Problems on Children's Problem Represenations and Solutions. *Journal of Educational Psychology*, 77(4), 460-470.

Desforges, A., & Desforges, C. (1980). Number-based Strategies of Sharing in Young Children. *Educational Studies*, **6**(2), 97-109.

Donaldson, M., & Balfour, G. (1968). Less is more: A study of language comprehension in children. *British Journal of Psychology*, **59**(4), 461-471.

Flavell, J.H. (1985). *Cognitive Development*. Englewood Cliffs, New Jersey: Prentice-Hall.

Frydman, O., & Bryant, P.E. (1988). Sharing and the understanding of number equivalence by young children. *Cognitive Development*, **3**, 323-339.

Fuson, K.C. (1982). An analysis of the counting-on solution procedure in addition. In Carpenter, T.P., Moser, J.M., & Romberg, T.A. (Eds.), *Addition and subtraction: A cognitive perspective*. Hillsdale, New Jersey: Lawrence Erlbaum Associates.

Fuson, K.C. (1988). *Children's Counting and Concepts of Number*. New York: Springer-Verlag.

Fuson, K.C., Secada, W.G., & Hall, J.W. (1983). Matching, Counting, and Conservation of Numerical Equivalence. *Child Development*, 54, 91-97.

Gelman, R. (1982). Accessing one-to-one correspondence: Still another paper about conservation. *British Journal of Psychology*, 73, 209-220.

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

Gelman, R., & Meck, E. (1983). Preschoolers' counting: Principles before skill. *Cognition*, 13, 343-359.

Gelman, R., Meck, E., & Merkin, S. (1986). Young Children's Numerical Competence. *Cognitive Development*, 1, 1-30.

Ginsburg, H.P. (1982). The development of addition in contexts of culture, social class, and race. In Carpenter, T.P., Moser, J.M., & Romberg, T.A. (Eds.), *Addition and subtraction: A cognitive perspective*. Hillsdale, New Jersey: Lawrence Erlbaum Associates.

Gréco, P. (1962). Quotité et quantité, nouvelles recherches sur la correspondence terme-à-terme et la conservation des ensembles. In Gréco, P., & Morf, A. (Eds.), *Strutures numériques elementaires*. Paris: Presses Universitaires de France.

Groen, G.J., & Parkman, J.M. (1972). A chronometric analysis of simple addition. *Psychological Review*, **79**, 329-343.

Groen, G.J., & Resnick, L.B. (1977). Can preschool children invent addition algorithms? *Journal of Educational Psychology*, **69**(6), 645-652.

Hiebert, J. (1981). *Young Children's Solution Processes for Verbal Addition and Subtraction Problems: The effect of the position of the unknown set*. Paper presented at the 59th Annual Meeting of the National Council of Teachers of Mathematics, St. Louis. Quoted in Briars, D.J., & Larkin, J.H. (1984). An Integrated Model of Skill in Solving Elementary Word Problems. *Cognition and Instruction*, 1(3), 245-297.

Hiebert, J. (1984). Children's Mathematics Learning: The Struggle to Link Form and Understanding. *Elementary School Journal*, **84**(5), 497-513.

Hitch, G., Cundick, J., Haughey, M., Pugh, R., & Wright, H. (1987). Aspects of counting in children's arithmetic. In Sloboda, J.A., & Rogers, D. (Eds.), *Cognitive processes in mathematics*. Oxford: Clarendon Press.

Hudson, L.M., Guthrie, K.H., & Santilli, N.R. (1982). The use of linguistic and non-linguistic strategies in kindergarteners' interpretations of "more and less". *Journal of Child Language*, **9**, 125-138.

Hudson, T. (1983). Correspondences and numerical differences between disjoint sets. *Child Development*, **54**, 84-90.

Hughes, M. (1986). *Children and Number - Difficulties in Learning Mathematics*. Oxford: Basil Blackwell Publications.

Ilg, F., & Ames, L.B. (1951). Developmental Trends in Arithmetic. *Journal of Genetic Psychology*, **79**, 3-28.

Inhelder, B., & Piaget, J. (1964). *The early growth of logic in the child*. London: Routledge & Kegan Paul.

Kintsch, W. (1974). *The representation of meaning in memory*. Hillsdale, New Jersey: Lawrence Erlbaum Associates.

Kintsch, W. (1988). The Role of Knowledge in Discourse Comprehension: A Construction-Integration Model. *Psychological Review*, 95(2), 163-182.

Kintsch, W., & Greeno, J.G. (1985). Understanding and solving word arithmetic problems. *Psychological Review*, 92, 109-129.

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

Littlefield, J., & Rieser, J.J. (1985). *Differentiation of Information: Locus of Gaze Patterns While Solving Math Story Problems*. Paper presented at the Biennial Meeting of the Society for Research in Child Development, Toronto.

Mayer, R.E. (1982). Memory for algebra story problems. *Journal of Educational Psychology*, 74, 199-216.

McShane, J. (1991). *Cognitive Development - An Information-Processing Approach*. Oxford: Basil Blackwell Publications.

Michie, S. (1984). Why preschoolers are reluctant to count spontaneously. *British Journal of Developmental Psychology*, 2, 347-358.

Moore, C., & Frye, D. (1986). Context, conservation and the meanings of more. *British Journal of Developmental Psychology*, 4, 169-178.

Moyer, J.C., Moyer, M.B., Sowder, L., & Threadgill-Sowder, J. (1984). Story Problem Formats: Verbal Versus Telegraphic. *Journal for Research in Mathematics Education*, 15(1), 64-68.

Muraski, V.S. (1978). *A study of the effects of explicit reading instruction on reading performance in mathematics and on problem-solving abilities of six graders*. Ph.D Thesis, Michigan State University. Quoted in Silver, E.A., & Thompson, A.G. (1984). Research Perspectives on Problem Solving in Elementary School Mathematics. *Elementary School Journal*, 84(5), 529-545.

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

Palermo, D.S. (1973). More about less: A study of language comprehension. *Journal of Verbal Learning and Verbal Behaviour*, 12, 211-221.

Piaget, J., & Inhelder, B. (1971). *Mental Imagery in the Child*. London: Routledge and Kegan Paul.

Piaget, J., & Szeminska, A. (1952). *The child's conception of number*. London: Routledge & Kegan Paul.

Pirie, S.E.B., & Kieren, T. (1989). A Recursive Theory of Mathematical Understanding. *For the Learning of Mathematics*, 9(3), 7-11.

Pirie, S.E.B., & Kieren, T. (1990). *A recursive theory for mathematical understanding - some elements and implications*. Paper presented at the AERA Boston Meeting, Massachusetts.

Pirie, S.E.B., & Kieren, T. (1991). *A Dynamic Theory of Mathematical Understanding - some features and implications*. Working paper presented at the AERA Chicago Meeting, Illinois.

Riley, M.S. (1981). *Conceptual and procedural knowledge in development*. Master's Thesis, University of Pittsburgh. Quoted in Riley, M.S., Greeno, J.G., & Heller, J.I. (1983). Development of children's problem-solving ability in arithmetic. In Ginsburg, H.P. (Ed.), *The development of mathematical thinking*. New York: Academic Press.

Riley, M.S., Greeno, J.G., & Heller, J.I. (1983). Development of children's problem-solving ability in arithmetic. In Ginsburg, H.P. (Ed.), *The development of mathematical thinking*. New York: Academic Press.

Saxe, G.B. (1977). A developmental analysis of notational counting. *Child Development*, 48, 1512-1520.

Saxe, G.B. (1979). Children's counting: The early formation of numerical symbols. *New Directions for Child Development*, 3, 73-84.

Schaeffer, B., Eggleston, V.H., & Scott, J.L. (1974). Number development in young children. *Cognitive Psychology*, 6, 357-379.

Siegler, R.S. (1987). Strategy choices in subtraction. In Sloboda, J.A., & Rogers, D. (Eds.), *Cognitive processes in mathematics*. Oxford: Clarendon Press.

Siegler, R.S., & Shrager, J. (1984). Strategy choices in addition and subtraction: How do children know what to do? In Sophian, C. (Ed.), *Origins of cognitive skills*. Hillsdale, New Jersey: Lawrence Erlbaum Associates.

Silver, E.A., & Thompson, A.G. (1984). Research Perspectives on Problem Solving in Elementary School Mathematics. *Elementary School Journal*, 84(5), 529-545.

Sophian, C. (1987). Early developments in children's use of counting to solve quantitative problems. *Cognition and Instruction*, 4, 61-90.

Starkey, P. (1983). *Some Precursors of Early Arithmetic Competencies*. Paper presented at the Biennial Meeting of the Society for Research in Child Development, Detroit, Michigan.

Starkey, P., & Cooper, R.G. (1980). Perception of numbers by human infants. *Science*, 210, 1033-1035.

Starkey, P., & Gelman, R. (1982). The development of addition and subtraction abilities prior to formal schooling in arithmetic. In Carpenter, T.P., Moser, J.M., & Romberg, T. A. (Eds.), *Addition and Subtraction: A Cognitive Perspective*. Hillsdale, New Jersey: Lawrence Erlbaum Associates.

Steffe, L.P., & Johnson, D.C. (1971). Problem-solving performances of first grade children. *Journal for Research in Mathematics Education*, 2, 50-64.

Tamburino, J.L. (1980). *An analysis of the modelling processes used by kindergarten children in solving simple addition and subtraction story problems*. Master's Thesis, University of Pittsburgh. Quoted in Briars, D.J., & Larkin, J.H. (1984). An Integrated Model of Skill in Solving Elementary Word Problems. *Cognition and Instruction*, 1(3), 245-297.

Vergnaud, G. (1982). A classification of cognitive tasks and operations of thought involved in addition and subtraction problems. In Carpenter, T., Moser, J., & Romberg, T. (Eds.), *Addition and subtraction: A cognitive perspective*. Hillsdale, New Jersey: Lawrence Erlbaum Associates.

Weaver, J.F. (1982). Interpretations of number operations and symbolic representations of addition and subtraction. In Carpenter, T.P., Moser, J.M., & Romberg, T.A. (Eds.), *Addition and Subtraction: A Cognitive Perspective. Hillsdale, New Jersey: Lawrence Erlbaum Associates.*

Weiner, S.L. (1974). On the development of "more" and "less". *Journal of Experimental Child Psychology*, 17, 271-287.

Whiteman, M., & Peisach, E. (1970). Perceptual and sensorimotor supports for conservation tasks. *Developmental Psychology*, 2, 247-256.

Wilkening, F., Levin, I., & Druyan, S. (1987). Children's Counting Strategies for Time Quantification and Integration. *Developmental Psychology*, 23(6), 823-831.

Woods, S.S., Resnick, L.B., & Groen, G.J. (1975). An experimental test of five process models for subtraction. *Journal of Educational Psychology*, 67, 17-21.

APPENDIX ONE

OTHER PROPERTIES OF WORD PROBLEMS

I. The Reading Aspects of Word Problems

II. The Relevant Versus Irrelevant Information Aspect of Word Problems

I. The Reading Aspects of Word Problems

Some studies suggest that reading ability plays only a minor role in problem solving performance. The National Assessment of Educational Progress (NAEP) in the United States conducted some mathematical assessments where children listened to word problems read on an audiotape. This was done in an effort to control the effect of reading difficulty. Even though the children did not have to read the problems, they still performed poorly on the problem tasks. On the other hand, other studies (Muraski,1978) have provided results which have demonstrated that teaching reading skills can lead to an increase in performance on word problem solving. Silver and Thompson (1984) [see Section 1.7.1.3] argued that it is overly simplistic to attribute a child's general lack of success in problem solving to reading difficulties.

One way of reducing the impact of reading ability is to use a "telegraphic" format where the verbal part of a word problem is greatly reduced (Moyer, Moyer, Sowder, and Threadgill-Sowder, 1984). Moyer et al. gave children (7-to-13-years-old) word problems presented in two formats: a verbal format and a telegraphic format.

An example of a word problem in the verbal format is the following:

Old funny books are 10 cents each.
Bill gets 3 old funny books. Marie gets 5 old funny books.
How much money will they pay for all the funny books?

An example of the word problem in its companion telegraphic format is the following:

Old funny books, 10 cents each.
Bill, 3 old funny books.
Marie, 5.
How much money will they pay for all?

The verbal and telegraphic problems were presented alternately to the children. Results demonstrated that at no age was the mean for the telegraphic format higher than the mean for the verbal format. Hence, Moyer et al. concluded that the telegraphic format did not facilitate children's performance on word problems. In fact, the successful problem-solver appeared to find the verbal format easier to interpret. The telegraphic format did not reduce the load on working memory. It is suggested that this may have been due to either of two possibilities: (1) the fact that the child was unfamiliar with the telegraphic format, or (2) the fact that the telegraphic format may not have corresponded to the way the child encoded the meaning of the sentences.

Moyer et al. considered the possibility that problem solving may require a first reading, where similar past situations are recalled, in order to choose schemata for solving the problem. This is followed by a second reading or a review of the data, where relevant information is collected. The verbal format, with its richer contextual information, may be more apt for the first part of the process. The telegraphic format, on the other hand, may be more apt for the second part of the process.

Though Moyer et al. were attributing their results to the reading aspects of word problems, their interpretation of the negative results may support a structural hypothesis. The possibility of a first reading followed by a second reading may be interpreted as a one-step and two-step process to arrive at the correct solution of a word problem.

II. The Relevant Versus Irrelevant Information Aspect of Word Problems

It has also been found that children find problems containing extraneous or superfluous data difficult to solve. There is a need to

distinguish relevant from irrelevant information. Children will have difficulty, for example, if they attempt to incorporate all the numbers in a word problem into the mathematical calculation, as some of these numbers may be irrelevant to the question.

Littlefield and Rieser (1985) also studied how mathematical story problems are solved. They argued that children of all ages are able to successfully solve problems that require a simple application of a single operation to all numbers given in the problem. They found that performance tended to be particularly bad when the problem required anything more than the direct application of a specifically taught algorithm, such as when irrelevant information was included in the problems.

Littlefield and Rieser thought that a distinction had to be made between relevant and irrelevant numbers in story problems and between relevant and irrelevant contextual information in story problems. They performed two studies with 10-and-11-year-old children. One study focused on children's differentiation of relevant from irrelevant numbers in story problems. The other study focused on children's differentiation of relevant from irrelevant contextual information in story problems. Each study presented the children with four story problems.

In both their studies, Littlefield and Rieser investigated three issues:

(1) Temporal changes in children's gaze patterns

These were investigated in order to see if they corresponded with the children's progress at differentiating relevant and irrelevant cues while working story problems. This was achieved by devising a simple method of recording locus of gaze.

(2) Differentiation patterns

These were investigated to see if they varied as a function of the child's mathematical ability. This was achieved by comparing the successful problem-solver's use of cues with those of the unsuccessful problem-solver.

(3) Differentiation strategies

These were investigated to see if they varied as a function of the type of irrelevant information present. This was achieved by comparing the child's differentiation of relevant numbers from irrelevant numbers, as well as the child's differentiation of relevant context from irrelevant context.

Hence, Littlefield and Rieser's research focused on the following measures: (1) locus of gaze, (2) percentage of problems solved correctly, and (3) time spent working on each problem. It was hypothesized that children who were successful in word problem solving would show better performance on the above-mentioned measures than less successful children.

Results demonstrated that problems containing irrelevant numbers elicited a lower percentage of correct answers and took longer to solve than problems with no irrelevant numbers or with irrelevant context. The locus of gaze data revealed the strategies children used to differentiate information. This type of data was analyzed in three ways:

(1) Statistical comparisons were made between the percentages of the total work time spent looking at each of the different categories of information.

(2) The data were grouped according to outcome of solution: [a] correct, [b] incorrect due to a computational error, or [c] incorrect due to the use of irrelevant information. Additionally they were grouped according to the number of gazes to relevant and irrelevant information in the first and second half of the work time.

(3) More qualitative analyses consisted of children's gaze patterns during problem solving in order to identify their search patterns.

More successful and less successful children did not differ with regards to the percentage of total time spent looking at problems with irrelevant numbers. Additionally, more successful children increased the time spent looking at relevant information and decreased the time spent looking at the workspace as problem complexity increased. Less successful children did

not change their pattern of gaze with problem complexity. As problem complexity increased for problems containing irrelevant context, neither the more successful nor the less successful children changed their patterns of gaze. Hence, Littlefield and Rieser's results appear to demonstrate that different types of irrelevant information have different effects on story problem performance of more and less mathematically successful children.

For problems that were solved correctly, or where only computational errors were made, children gazed more at relevant than irrelevant information. In the case of problems that were solved incorrectly due to the use of irrelevant information, children gazed more at relevant than irrelevant information in the first half of the solution time, but not in the second. Children were more successful at differentiating irrelevant contextual information from irrelevant numerical information, and were, therefore, more likely to solve such problems correctly.

The possibility exists that the amount of irrelevant numbers and irrelevant context in a word problem affects the number of steps necessary to arrive at the correct solution of the problem. Hence, though Littlefield and Rieser were attributing their results to the relevant versus irrelevant information aspect of word problems, support for a structural aspect is also found.

From a different perspective, Nesher and Teubal (1975) investigated the relevant versus irrelevant information aspect of word problems. They looked at children's transition from natural language to the corresponding mathematical expression and distinguished three levels. The first level was the verbal formulation level, the second level was the underlying mathematical relations level, and the third level was the symbolic mathematical expression level. According to Nesher and Teubal, transition can be made from Level 1 to Level 2 to Level 3, as well as from Level 1 to Level 3.

Nesher and Teubal designed an experiment to test the hypothesis that there are key words in the verbal formulation of mathematical problems which affect the choice of a mathematical operation. These key words can either be relevant or irrelevant to the mathematical operation. Nesher and Teubal thought that if a child is confronted with a problem where either addition or subtraction could be used to obtain a correct solution to the problem, then the choice of adding or subtracting will depend on the key words involved in the problem formulation. To test their hypothesis, Nesher and Teubal presented story problems to the children, where either adding or subtracting could be used to attain the correct solution.

Results demonstrated that children preferred addition to subtraction. Furthermore, the choice of the addition operation was dependent on the key word (the word "buying", for example) in the problem. The choice of the subtraction operation, however, was not dependent on the key word (the word "losing", for example) in the problem. Nesher and Teubal concluded that certain key words do have a definite influence in determining the child's choice of mathematical operation.

In another experiment, Nesher and Teubal presented a group of children with four types of problem: (1) those with the comparative term "more" which required the addition operation, (2) those with the comparative term "more" which required the subtraction operation, (3) those with the comparative term "less" which required the subtraction operation, and (4) those with the comparative term "less" which required the addition operation. Results demonstrated that the largest proportion of correct responses occurred when the key words acted as "clues" to the operation. Results also demonstrated that the proportion of correct responses diminished when the key words appeared as "distractors" from the operation.

Nesher and Teubal concluded that children failed to use Level 2, the underlying mathematical relations level, as a transition between Level 1, the

verbal formulation level, and Level 3, the symbolic mathematical expression level. Children seemed to be going from Level 1 to Level 3.

Thus, according to Nesher and Teubal, successful problem-solving may be due to either a successful interpretation or translation of the problem or a successful dealing with more specific discrete structural values. Some linguistic aspects, they suggest, may cause confusion. Certainly these more specific discrete values must be more evident in some types of word problem than in others. Their bearing upon children's performance on Equalize and Compare word problems should be considered. These experimenters are one step closer at considering the difference between Equalize and Compare word problems.

APPENDIX TWO

ADDITIONAL WORD PROBLEM MODELS

I. The Kintsch Model
II. The Langford Model
III. A Further View of the DeCorte and Verschaffel Model

I. The Kintsch Model

Kintsch's (1988) Construction-Integration Model of word problem solving is based on discourse comprehension. His model deals with concepts and propositions being formed, inferences and elaborations being produced, and word meanings being activated. These form a connectionist network of interrelated elements which are integrated into a coherent structure through a spreading activation process. This spreading activation process quickly takes care of inconsistencies and irrelevancies. Flexibility and context sensitivity result from this.

According to Kintsch, construction processes involve: (1) forming the concepts and propositions that directly correspond to the linguistic input, (2) elaborating each of these elements by selecting a small number of its most closely associated neighbours from the general knowledge net, (3) inferring additional propositions, and finally, (4) assigning connection strengths to all pairs of elements that have been created. Following all these steps of the Construction-Integration Model, a child will eventually arrive at the correct interpretation of the problem.

According to Kintsch, the ambiguity characteristic of language accounts for many misinterpretations and errors to elicited questions. Hence, "general knowledge" about words, syntax, the world, spatial relations, etc., sets constraints on the construction of discourse representations at all levels. However, these constraints are what make it possible to build the representations. Therefore, "general knowledge" makes understanding possible and correct understanding depends on knowing what to expect.

It is not possible to deal with the whole, general knowledge net at once. Only tiny parts of the net can be activated, and only those propositions of the net that are actually activated can affect the meaning of a given concept. Hence, the meaning of a concept or word is always dependent on the situation and on the context.

Solution to mathematical word problems can either be right or wrong. "Knowledge" which needs to be activated in the mathematical operation of word problems is pretty much well defined and domain-specific. Ambiguous criteria for understanding should not exist. However, according to Kintsch's model, the meaning of words in a discourse may facilitate or interfere with the understanding of the problem. Word identification is not simply a matter of lexical access, but a complex process that responds to different influences at different stages.

Context effects can not be overlooked. They are expectation driven and may facilitate or interfere with the perceptual analysis. "Association" means that several meanings are attributed to a word, and perceptual analysis alone cannot decide which meaning of a word to choose for a given context. Hence, the meaning of a word is discovered by "association" and the meaning that fits depends on the context. Kintsch labelled this the "sense-selection" stage. Kintsch also referred to a "sense elaboration phase", where as more information about the context becomes available, the sentence and discourse meaning begin to emerge. The meaning of a word is explored and elaborated contextually. Word meanings are usually identified long before complex inferences are made in comprehending a discourse.

Henceforth, a word is perceived in a context-dependent way. According to Kintsch, it activates its whole associative neighbourhood. Consequently, strong associates of a word are likely to be represented in working memory and therefore will be primed in a lexical decision task, whether they are context appropriate or not. The knowledge-integration process then results in the deactivation of material that is irrelevant to the overall discourse context, such as the context-inappropriate associates.

According to Kintsch, mathematical knowledge forms a special subset of a person's general knowledge network. Sets of objects can be represented by a propositional schema with the following components: object,

specification, quantity, and role (relationship of sets of objects to other sets). This propositional schema gives way to a superordinate schema where there is a transfer-in schema with the following components: a start set, a transfer-in set, and a result set. This superordinate schema, according to Kintsch and Greeno (1985) comprises various mathematical strategies, such as counting strategies. Appropriate mathematical strategies generate the right solutions when solving word problems. There are three forms of mathematical strategies:

(1) The strategies that are based on the propositional schema.

(2) The strategies that determine the nature of the connection between the propositional schema and the mathematical hypotheses.

Whenever a proposition is encountered in the text base, possible mathematical hypotheses derivable from it are constructed. These hypotheses may involve two identical propositions with the roles "whole" and "part". Key words connect propositions in the text base to the hypotheses. Collection terms, such as "altogether", indicate whole sets, whereas "give/take", "of these", and "have more/less than" indicate part sets. What is "whole" and what is "part" is determined by general knowledge about situations and actions.

Kintsch (1984) had previously described certain aspects of the strategies that determine the nature of the connection between the propositional schema and the hypotheses. These are:

(a) Restricted Subsets

If the specification of one set is more general than that of another, then the former is assigned the role of whole and the latter is assigned the role of part.

An example of this is the following:

large-window, small-window versus *window*

on-upper shelf, on-lower shelf versus *on-shelf*

(b) Conjunction

If the specification of one set consists of the conjunction of the specification of two other sets, then the former is assigned the role of whole and the others the role of parts.

An example of this if the following:

yesterday, today, and *yesterday&today*

teddy bear, doll, and *toy*

(c) Time-ordered possession/location

If the specification of three sets contain either (1) HAVE (possession), (2) LOCATION (place), (3) the negation of these propositions, or (4) information to establish temporal order, then WHOLE and PART roles can be assigned to the three sets according to the resulting patterns.

(3) The strategies that form the superordinate schema on the basis of the mathematical operations.

This concerns the part-whole schema and depends on whether it is the first, second, or third set which is considered the whole set. The order of the sets in the word problem and not the temporal order is of issue here.

An example of this is the following:

A part-part-whole problem where the construction of that problem is a part-whole one with the third set as the whole:

PPW[role[set1,part],role[set2,part],role[set3,whole]]

Kintsch believed that his model can account for some well known facts which Briars and Larkin's model (1984) and Kintsch and Greeno's model (1985) can not. Kintsch's model (1988) can account for:

(1) <u>Inferences</u>

"Manolita tried to weed her father's garden. 'You sure weeded it', said Mr. Mundoza. 'There were fourteen tulips in the garden and now there are only six.' How many tulips did she pull out by mistake?"

In order for children to solve this problem, they need to apply one of the location strategies: there were X amount of tulips, some were pulled out, and so many are left. A simple knowledge-based inference is necessary here.

In the above-mentioned example, the tulips that were pulled out were no longer in the garden. If the knowledge-activation mechanism supplies the necessary inference, then the problem will be solved successfully. Kintsch's model will solve this problem in three cycles: (a) the first sentence which only really sets up a context and is not really relevant to the mathematical operation, (b) Mr. Mundoza's statement, and (c) the question sentence.

It must be inferred that the 14 tulips that were in the garden in the past are, a part or whole set, and the six tulips now in the garden are a part or whole set. The tulips before the garden was weeded are the whole set, and only a part is left after the weeding. Hence, past is connected with whole and present with part. At this point, the word problem is actually understood correctly and practically solved as "whole", which is more strongly activated than its alternative "part". "Whole" is stronger than "part". The text propositions and the inferences are more strongly activated than the mathematical hypothesis. Hence, the mathematical hypothesis has to be strongly anchored in a stable text representation. This implies that the language used is very important.

The understanding of the problem is then completed by establishing textual coherence from the text propositions. The location strategy completes the pattern. (Some tulips were at one place in the past, then some were not there, now some are left.) Here, the knowledge-based inference achieves an activation level above the range of the text propositions. The Manolita problem was in fact solved without problem solving. The basic text propositions were sufficient to produce the inference that the pulled-out tulips were not in the garden, which was required for the application of the location strategy. However, the Manolita problem is not considered a difficult problem. For other more difficult problems, the random inference generation process, described by Kintsch, fails to generate the required inference.

(2) <u>Context Effects</u>

Problems embedded into a familiar situational context are much easier to solve than problems that must be solved without this situational support (Hudson,1983) [see Section 1.7.2]. In Hudson's problems, birds catching worms represent a situation which is not only understandable, but concrete. Additionally, the part-whole relationship is clear. According to Kintsch, the situations involved in Hudson's problems have little to do with mathematics and it is the knowledge about birds eating worms that matters. The birds trying to catch the worm are understood as the whole set, the bird catching worms as one part and the birds unable to get a worm as the other part. In Kintsch's view, the problem is understood on the basis of "general knowledge", and not because an understanding of the question, *"How many more...?"*, is achieved. Kintsch states that the mathematical operation is unlikely to go wrong here because the situation is presented in such a way that only the right interpretation of the problem can occur. The right interpretation of the problem is guaranteed by the well-established situation.

Context does not always assist in the solution of a problem. It sometimes can interfere with it. Kintsch provided an example where context is irrelevant to the appropriate solution of the problem. The following problem must be solved with specialized mathematical strategies using the help of the key words available, like "have altogether" for the first question, and "have more than" for the second question.

Fred has four Chevies and three Fords.
(1) How many cars does he have altogether?
(2) How many more Chevies does he have than Fords?

(3) <u>Question Specificity</u>

Overspecifying can be useful as it provides more than one way to tackle a problem. However, looking at this from a different perspective, redundant specifications increase the length of the text and thus, the likelihood that some crucial bit of information is no longer considered relevant.

Hence, Kintsch's model of discourse comprehension has two stages: (a) a propositional network that must be constructed and, (b) that must be edited or integrated. The words and phrases that make up the discourse are the raw material from which a mental representation of the meaning of that discourse is constructed. The mental representation takes the form of a propositional text base. This text base combines information which is in the text itself and in the knowledge possessed by the person dealing with the text. This knowledge can be knowledge about language, as well as knowledge about the world.

"General knowledge" used in discourse comprehension in Kintsch's model has two levels: one is a general computational mechanism and the other specifies how this mechanism is used in word identification, in discourse, and in understanding and solving word problems. Thus, the model

also helps in detecting which formulation of a problem is easy and which formulation of a problem is difficult.

II. The Langford Model

Langford (1986) also sought to create a model of word problem solving. He attempted to gain further information concerning strategies and error types of 6-year-old children's performance on word problems using concrete materials.

Langford studied 40 children (20 girls and 20 boys) ranging in ages between 6-to-7-years-old. He gave the children a slightly modified sample of the questions used by Riley et al. (1983) under two conditions. Children in both conditions had marbles, which they could use in modeling the word problems. In the first condition, the children were read the instructions twice and were then allowed to begin modeling. In the second condition, the children were read the instructions once and were then allowed to request any further information that they desired. Half of the children were presented with Condition 1 first and Condition 2 second, and the other half of the children were presented with Condition 2 first and Condition 1 second. The questions were asked in the same order for both conditions. Questions in their respective order for Condition 1 are detailed below. For Condition 2 the questions were the same but with different numbers such that neither the numbers in the question nor in the correct answer exceeded 9.

Change 1: **"You have 5 marbles. Then I give you 4 more marbles. How many marbles do you have now?"**

Change 3: **"You have 4 marbles. Then I give you some more marbles. Now you have 7 marbles. How many marbles did I give you?"**

Change 5: **"You have some marbles. Then I give you 4 more marbles. Now you have 6 marbles. How many marbles did you have in the beginning?"**

Combine 2: **"You have 4 marbles. I have 3 marbles. How many marbles do we have altogether?"**

Change 2: **"You and I have 9 marbles altogether. You have 4 marbles. How many marbles do I have?"**

Compare 1: **"You have 7 marbles. I have 4 marbles. How many marbles do you have more than me?"**

Compare 3: **"You have 4 marbles. I have 3 more marbles than you. How many marbles do I have?"**

Compare 5: **"You have 9 marbles. You have 4 more marbles than me. How many marbles do I have?"**

In Condition 2, most of the children chose concurrent modeling. Moreover, the questions most prone to linguistic misinterpretation, under both of the conditions were Change 3 and 5, Combine 2, and Compare 1 and 3. For example, *"How many marbles did I give you?"*, was often misinterpreted as, *"How many marbles did you have?"*; *"You and I have 9 marbles altogether"*, was often misinterpreted as, *"You and I each have 9 marbles"*; *"How many marbles did you have more than me?"*, was often misinterpreted as, *"How many marbles did you have?"*; *"I have 3 more marbles than you."*, was often misinterpreted as, *"I have 3 marbles."*. Langford points out that these types of error are commonly confused with random guessing. The

expected number of errors in this Guessed a Question Value Strategy was low except for Change 3 and 5 under Condition 1, where this type of error occurred. Linguistic error was very high for Compare 1 in Condition 2.

Langford observed another strategy known as Transformed Unknown to Concrete Value Strategy. An unknown is mentioned in the problem and the child either guesses a value or inserts a zero as the value in order to proceed with model building. Briars and Larkin (1984) [see Section 1.7.1.2] had suggested that this problem is typically resolved by transforming the question into a form that has a single unknown on one side of the equation and avoids any unknowns on the other side. This latter part of the question can then be modelled without unknowns. Hence, *"You have 4 marbles. Then I give you some more marbles. Now you have 7 marbles. How many marbles did I give you?"*, is transformed into, *"The same marbles given equal the 7 marbles you have at the end minus the 4 marbles you had at the beginning"*. This is then modelled by counting out 7 and removing 4. Langford discovered, however, that successful problem solvers did not proceed in this way. Most of the successful problem solvers began with 4, incremented to 7 and then looked at the increment. They lost track of the incremented collection by failing to physically separate it from the original collection and hence, went wrong in their attempt to solve the word problem. This strategy and its common substitution of concrete values for the unknown fits better with Langford's claim that children solving this problem are able to consider unknown sets during the process of model building. Question types Compare 3 and 5 involve consideration of unknown sets but no Transformed Unknown to Concrete Value Strategy was evident.

Condition 2 involved the conventional way of asking questions. The distribution of error types and strategies used in order to arrive at the correct solution were very similar to those of Condition 1. Langford suggested that Condition 1 did not force children to adopt fundamentally different strategies

for problem solving; instead it seemed that the information from the questions was stored into short-term memory, as Kintsch and Greeno (1985) suggested. According to Langford, the children used this short-term memory for problem solving at a time when in the strategy adopted for Condition 2 they would have turned to the experimenter for more direction.

Results showed that Condition 1 encouraged more correct strategies without modeling. There was a significant tendency to convert the question directly into a mathematical form without using modeling in order to arrive at the solution. Of the 20 "wrong strategy" cases, 18 of them involved the use of modeling. According to Langford, this provides evidence for behaviour of the type modelled in the Riley et al. (1983) [see Section 1.7.1.1] and Kintsch and Greeno (1985) models. In these models an internal schema is used to solve the problem and modeling is used as a method for computing the answer. The possibility exists that a proportion of the correctly modelled solutions in this condition was produced in this fashion.

Langford compared his results for Condition 2 with those of Riley (1981) [see Section 1.6.1]. The Combine question in Riley's study was solved correctly by 39% of the children, whereas in Langford's study, it was solved correctly by 90% of the children. Compare question 3 in Riley's study was solved correctly by 17% of the children, whereas in Langford's study it was solved correctly by 50% of the children. Proportions in Langford's study were altered by eliminating the non-concurrent modelers and by treating slips in counting and computational errors as correct. This, however, does not explain the discrepancies, which are evident in other proportions, in Riley's (1981) study. DeCorte et al. (1985) had shown that minor changes in the wording can considerably influence the difficulty of mathematical word problems. Langford suggests that children may have performed better in his study because Riley's "Joe" and "Tom" were changed to "You" and "I". This form of presentation was easier for the children.

Hence, Langford considered that the children's relative difficulty in word problem solving is affected by minor changes in wording and that some questions give a considerable rise to a substantial proportion of linguistic errors. According to Langford, any hypothesis that does not take into account the influence of semantic comprehension on the problem's level of difficulty is not to be taken seriously.

III. A Further View of the DeCorte and Verschaffel Model

DeCorte and Verschaffel (1987b) constructed a model in an attempt to explain the difficulty children have when solving Compare word problems. Their model is based on their evidence, which has previously been presented (see Section 1.6.1), for the difficulty children have with Compare problems. This model can be divided into five stages.:

(1) Text processing action commences which is fairly complex and goal-oriented. Starting from the verbal text, the child builds an internal representation of the problem in terms of sets and set relations. This internal representation is global and abstract. This building of an internal representation is considered to be the most crucial component of skilled problem solving. There are two factors which influence the construction of this internal representation: (a) the representation of the child's knowledge about increasing and decreasing, combining, and comparing groups of objects, known as the three components of semantic schemata: Change, Combine, and Compare; and (b) the knowledge of the structure of word problems, their role and intent in mathematical problems, and the implicit rules and assumptions underlying that particular type of text, known as the word problems schema.

(2) From this representation, the problem-solver then chooses an appropriate mathematical operation or an informal counting strategy in order to find the unknown element in the problem representation.

(3) The selected action or operation is executed.

(4) Reactivation of the initial problem representation is done by the word problem-solver by first replacing the unknown element with the result of the action performed and then by formulating the answer.

(5) Verification actions are performed to double check the correctness of the solution as found when selecting the action or operation.

This model can explain the difference between Equalize and Compare word problems. Equalize word problems are represented by an execution of the first three stages of the model, whereas Compare word problems are represented by an execution of the first four stages of the model. Hence, in order for children to successfully perform on Compare word problems, they need to arrive one step further from Equalize word problems, in DeCorte and Verschaffel's model. This model then supports the one-step/multi-step structural hypothesis explaining the Equalize/Compare difference (see Section 1.7.1.3). Stage (4) in this model, also supports the parallel drawn to work on causal reasoning by DasGupta and Bryant (1989) [see Section 1.7.1.3], as one has to refer back to (or reactivate, according to DeCorte and Verschaffel) the initial state (or representation, according to DeCorte and Verschaffel) of the problem. Furthermore, this model strongly supports the model put forth by this author [See Section 1.7.3].

DeCorte and Verschaffel (1987b) argue that a well developed representation of a word problem is not only the result of the semantic processing of the verbal text but also of a pragmatic interpretation process which is facilitated by the command of the word problem schema. Some children, however, interpret the semantic representation from the verbal text inappropriately and hence, make errors.

The DeCorte and Verschaffel model was directly influenced by the model of Riley, Greeno, and Heller (1983) [see Section 1.7.1.1]. However, the DeCorte and Verschaffel model expands the latter. Like Riley et al., DeCorte and Verschaffel believe that the construction of an appropriate representation of the problem is crucial for successful problem solving. Unlike Riley et al., DeCorte and Verschaffel further add that children's difficulty with certain type word problems is also due to a misunderstanding of isolated words and/or sentences in the verbal text, (i.e. word problem schema), which leads them to an initially incorrect text base.

DeCorte and Verschaffel (1987a) had found that the development of the word problem schema was directly and significantly related to problem solving ability or the child's capacity to work out the difference between two unequal quantities on simple addition and subtraction word problems. DeCorte and Verschaffel also examined 5-year-olds' errors which did not result from a fault in their word problem schema, but which were a result of an incorrect semantic representation of the whole or of parts of the verbal text. This, they found, affects three types of problems: Compare, Change, and Combine problems. It has previously been shown (Hudson, 1983) [see Section 1.7.2] that children when solving such problems incorrectly, often reproduce as their response, one of the numbers given in the text of the problem.

Children's difficulty in solving these Compare-type problems had been attributed to a lack of procedural knowledge (Carpenter, Hiebert, and Moser,1981; Riley et al.,1983). This procedural knowledge enables the child to find out the difference between the two sets that are being compared. However, Riley et al. (1983) do not believe this to be the full explanation. They believe that children failed to solve the Compare-type problems as a result of an inappropriate representation of the word problem, due to the

absence of a schema which allows the child to understand the problem situation.

DeCorte and Verschaffel claimed that children interpreted a relation proposition, as assignment propositions in Compare-type problems. A relation proposition is a term taken from Mayer (1982) and is a proposition that involves drawing a single numerical relationship between two variables. This relation proposition may be interpreted as assignment propositions, which are propositions assigning a single numerical value for some particular variable. For example, the proposition, *"Ann has 6 more apples than Pete."*, is interpreted as, *"Ann has 6 apples."*, and the question, *"How many more apples does Ann have than Pete has?"*, is interpreted as, *"How many apples does Ann have?"*. This was true both of the retelling task, and of the material representation task where children represented the problems with puppets and blocks.

DeCorte and Verschaffel emphasize that some of the children's errors are due not only to their inability in building a cohesive problem representation on the basis of the propositional text base, but also to a misunderstanding of isolated words and/or sentences in the verbal text which leads to an initially incorrect text base. According to them, errors could also be due to a lack of a well-developed word problem schema since they do not have any familiarity with the task. Hence, not all errors are attributed to difficulties with the semantic processing.

There are certain task characteristics in the initial stage where construction of the appropriate problem representation takes place which definitely affect the level of difficulty of the elementary addition and subtraction problems. These task characteristics affect the nature of the semantic relationship between the known and unknown quantities in the problem, such as the cause-Change, the Combine, and the Compare relationships. Another task characteristic is the nature of the unknown set

(whether it is in one of three positions: the start set, the change set, or the end set). A third task characteristic is the degree in which the underlying semantic relations between the quantities are stated explicitly in the problem text.

Though DeCorte and Verschaffel did not deal with Equalize word problems, their model could explain the difference between Compare and Equalize problems. 1) The task characteristics, which affect the nature of the semantic relationship between the known and unknown quantities, differ between Compare and Equalize problems. A Compare problem, as seen previously, has a semantic relationship, such as 6-4=X, whereas an Equalize problem, has a semantic relationship, such as 4+X=6. 2) Another task characteristic, which also affects the nature of the semantic relationship is the position of the unknown set. This, also varies between Compare and Equalize problems. The unknown in a Compare problem is usually the end set, whereas in an Equalize problem it is usually the change set. 3) Finally, Compare and Equalize problems differ in the degree in which the underlying semantic relations between the quantities are stated explicitly in the problem text. Compare and Equalize problems vary in their semantic construction.

In general, DeCorte and Verschaffel have two main conclusions. Based on their studies, they point out that children have, at least implicitly, a quite sophisticated knowledge of number and elementary operations such as addition and subtraction and they also seem to apply this knowledge to a variety of situations. A second main conclusion concerned children's performances when solving word problems differing with respect to semantic structure that could be solved by the same operation. These, they note, can have different degrees of difficulty. Evidence for the influence of semantic structure on problem difficulty is also found in other studies (Hudson,1983).

Thus, according to DeCorte and Verschaffel, children's performance on word problems depends on the degree of difficulty of the story presented.

According to these experimenters, the way that children perform in Compare and Equalize problems depends on the semantic structure of these problems.

Though DeCorte and Verschaffel's model emphasized word problem differences on a combination of structural and linguistic aspects in a numerical context, it is speculated that their model would also apply to word problem differences in a non-numerical context. Hence, it is speculated that this model would hold valid, not just in its numerical context, but also in a quantitative context, as it is hypothesized that it is not only specific to counting.

Should this model apply to both continuous and discontinuous quantity, then, without a need for any further evidence, it can be claimed that the Equalize/Compare difference refers not only to number, but more fundamentally, to quantity in general.

APPENDIX THREE

Test devised for Experiment 1

Fill in the blanks with the correct answer:

1. **3 + 4 = _____**

2. **6 - 2 = _____**

3. **5 + 2 = _____**

4. **7 - 6 = _____**

5. **3 + 5 = _____**

6. **7 - 4 = _____**

7. **3 + 3 = _____**

8. **8 - 4 = _____**

APPENDIX FOUR

CHILDREN'S ABILITY TO REPRESENT ONE QUANTITY

APPENDIX FOUR

CHILDREN'S ABILITY TO REPRESENT ONE QUANTITY

Appendix 4 **Experiment 5A** **An Investigation of Responses made in Different Types of Displays using Different Types of Continuous Material**

5A.1 Introduction

Previous data (Experiment 5) demonstrated that spatial imagery plays some part in the understanding of both Equalize and Compare questions. This was supported by the highly significant independent effect of condition in Experiment 5, which demonstrated that children find same-type material displays easier than different-type material displays. However, as there was no interaction between spatial imagery and question type, this finding does not explain the Equalize and Compare difference, as it applies to both these question types.

Before assuming that same level displays and different level displays would have an effect on children's performance on Equalize questions, a preliminary experiment is needed on length comparison. Experiment 5A is designed to introduce these different levels and investigate whether children can compare length easier when it is presented at a same level than when it is presented at a different level with different type of material displays. A different type material, other than bars and lines, such as men, was introduced. (See Figure 5A.1.1.) Children may be able to easily envisage increasing or decreasing a single quantity. Hence, by providing men, children should have no problem envisaging a man growing or a man shrinking.

5A.2 Method

(a) Subjects

Two groups of children served as subjects: one consisting of 21 children ranging in age from 6;3 - 7;3 (mean age of 6;9), and the other of 22 children, ranging in age from 7;4 - 8;1 (mean age of 7;6). The children were all attending a first school in Oxford, England.

(b) Apparatus

The same apparatus was used as for Experiments 2, 3, and 4.

(c) Design

Each child was given one sequence, in which there were 36 trials, where the child had to choose out of five choice stimuli, the same one as the comparative stimulus. Eighteen of these 36 trials were for the same level display and the other 18 trials were for the different level display. A same level display constituted the comparative and choice stimuli spatially arranged to begin at the same base level. A different level display constituted the comparative and choice stimuli spatially arranged to begin at different base levels; in this spatial arrangement the comparative stimuli was higher than the choice stimuli. The top of the comparative stimuli was raised to the height of the highest item of the choice stimuli. (See Figure 5A.1.1.) Each of the 36 trials was blocked into six conditions. There were six sequences of the six conditions, the order of which was counterbalanced. The six sequences were the following:

Sequence1	Sequence2	Sequence3	Sequence4	Sequence5	Sequence6
MM	BB	LL	MB	LB	ML
BB	LL	MB	LB	ML	MM
LL	MB	LB	ML	MM	BB
MB	LB	ML	MM	BB	LL
LB	ML	MM	BB	LL	MB
ML	MM	BB	LL	MB	LB

where the material conditions were (see Figure 5A.1.1):

(1) Condition MM. The comparative stimuli and the choice stimuli were men. This was a same material condition.

(2) Condition BB. The comparative stimuli and the choice stimuli were bars. This was a same material condition.

(3) Condition LL.The comparative stimuli and the choice stimuli were lines. This was a same material condition.

(4) Condition MB. The comparative stimuli was a man and the choice stimuli were bars. This was a different material condition.

(5) Condition LB. The comparative stimuli was a line and the choice stimuli were bars. This was a different material condition.

(6) Condition ML. The comparative stimuli was a man and the choice stimuli were lines. This was a different material condition.

The display variable was presented in alternating order. For both groups of children, half of the subjects received the same level display first followed by the different level display (sequence-set A); the other half were given the task in the reverse order (sequence-set B). The design for Experiment 5A is illustrated in Table 5A.2.1.

(d) **Materials**

The continuous computer-driven displays used are presented in Figure 5A.1.1. There were six displays (three in a same level display and an equivalent three in a different level display) for each of the six material conditions, described in the design section (c) of this experiment.

I. Same-type Material Display

This type of display constituted the comparative stimulus and the choice stimuli to be composed of the same display.

A. Man-Man Display

Comparative Stimuli

On the screen, six stimuli were presented. The comparative stimulus consisted of one man. It was selected from a set of 10 men, which were 1,2,3,4,5,6,7,8,9, or 10 inches high. In two displays, it was either 1,2,3,4,5,6,7, or 8 inches high; in the other display, it was either 9 or 10 inches high. The comparative stimuli was always the same height as one of the choice stimuli.

Choice Stimuli

The choice stimuli consisted of five men which were 1,2,3,4,5,6,7,8,9,10,11,12,13, or 14 inches high.

In nine displays, the choice stimuli were arranged in ascending order of height, while in the other nine displays they were arranged in descending order. The choice stimuli that were arranged in ascending order in nine of the displays were as follows:

Comparative Stimuli	Choice Stimuli
1,1	2 displays that were 1,3,5,7,and 9 inches in height
2,2,4,4,8	5 displays that were 2,4,6,8,and 10 inches in height
7,7	2 displays that were 3,5,7,9,and 11 inches in height

The choice stimuli that were arranged in descending order in nine of the displays were as follows:

Comparative Stimuli	Choice Stimuli
8	1 set of items that was 10, 8, 6, 4, and 2 inches in height
3,3	2 sets of items that were 11, 9, 7, 5, and 3 inches in height
6,6	2 sets of items that were 12,10, 8, 6, and 4 inches in height
5,5,9	3 sets of items that were 13,11, 9, 7, and 5 inches in height
10	1 set of items that was 14,12,10, 8, and 6 inches in height

Hence, in half of the displays, the numbers of inches in height of the items were all odd, while in the other half they were all even.

In four displays, the correct response was in the first position of the choice stimuli; in three, it was in the second position of the choice stimuli; in four, it was in the third position of the choice stimuli; in three, it was in the fourth position of the choice stimuli; and in four it was in the fifth position of the choice stimuli.

In all displays, the items (men) in each set were positioned one inch apart, and there was a distance of two inches between the two stimuli sets.

Any display in any condition could be blue, pink, yellow, or green in colour, and the colour of any given display was selected randomly.

B. Bar-Bar Displays

The Bar-Bar Displays were similar to the Man-Man displays, except that the quantities were here represented by bars. The same design was used as for the Man-Man displays, except that, for example, a 7-inch man was replaced by a 7-inch bar. (See Figure 5A.1.1.)

C. Line-Line Displays

The Line-Line Displays were similar to the Man-Man displays, except that the quantities were here represented by lines. The same design was used as for the Man-Man displays, except that, for example, a 7-inch man was replaced by a 7-inch line. (See Figure 5A.1.1.)

II. Different-type Material Display

D. Man-Bar Displays

The Man-Bar Displays were similar to the Man-Man and Bar-Bar displays, except that the quantities were represented by both a man and bars. The same design was used as for the Man-Man and Bar-Bar displays, except that, the comparative stimulus was here represented by a man and the choice stimuli by bars. (See Figure 5A.1.1.)

E. Line-Bar Displays

The Line-Bar Displays were similar to the Bar-Bar and Line-Line displays, except that the quantities were represented by both a bar and lines. The same design was used as for the Bar-Bar and Line-Line displays, except that, the comparative stimulus was here represented by a line and the choice stimuli by bars. (See Figure 5A.1.1.)

F. Man-Line Displays

The Man-Line Displays were similar to the Man-Man and Line-Line displays, except that the quantities were represented by both a man and lines. The same design was used as for the Man-Man and Line-Line displays, except that, the comparative stimulus was here represented by a man and the choice stimuli by lines. (See Figure 5A.1.1.)

The differences between Experiment 5A from Experiments 2, 3, and 4 can be summarized by the following points:

(1) The comparative stimuli was composed of one item. Depending on the display, the comparative stimuli consisted of either (a) one man, (b) one bar, or (c) one line.

(2) The display variable (same level or different level) was composed of 3 material conditions (men, bars, or lines).

(e) Procedure

Children were taken from their classes individually, in a random order. The children were shown each display one at a time as presented on the screen. Each of the 18 displays was presented twice under each of two display types (same level or different level).

Children were asked one question for each display. Hence, each child was asked six questions per material display (Condition MM, Condition BB, Condition LL, Condition MB, Condition LB, Condition ML), totalling 36 questions in all. The question asked was the same for each display.

The question took the form: *"Which one of these* (the experimenter pointed to the comparative stimuli) *is the same as this one* (pointing to the choice stimuli)*?"*. The child then selected an answer (one of the five men, bars, or lines of the choice stimuli). Responses were recorded and stored in the computer.

5A.3 Results

The data were the number of questions answered correctly. The mean number of correct responses for each type of condition in each display type is shown in Table 5A.3.1. These means were obtained from scores on the 6

questions on each of the six types of material display and give an overall picture of children's performance.

It was predicted that children would find the different level displays particularly difficult when presented with different-type material. In general, more correct responses occurred in the same level display than in the different level display. As predicted, children seemed to find different level displays difficult when presented with the different-type material.

Table 5A.3.1 **Means and Standard Deviations (out of 3) for the Number of Total Correct Responses**

5A.3.1 Comparing Correct Scores across Conditions

The means and standard deviations for the total number of correct responses made in each display for each material condition by the children were given in Table 5A.3.1. This table shows that both age groups elicited about the same amount of correct responses. Same level displays were superior to different level displays. Same type of material display yielded more correct responses from the children than different type of material display. Thus, children performed better in the MM, BB, and LL conditions than in the MB, LB, and ML conditions.

Calculating the Probability of Correct Responses occurring purely by Chance

As for Experiments 2, 3, and 4, children had a choice of five responses (from the choice stimuli). Therefore, there was a 20% chance of their choosing any one of these responses. Children were considered to be performing at chance level if they got a total of .6 correct responses. The probability that the correct responses occurred simply by chance was

explored using a modification of the standard hypothesis testing procedure with the means in Table 5A.3.1. (See Section 3.3.1.)

The proportion of correct responses made in all cell means were significantly ($p<0.01$) greater than expected by chance. This indicates that the children were not acting randomly.

5A.3.2 Analysis of Variance

These trends were analyzed by subjecting the raw scores for the total number of correct responses to a 2x2x6 ANOVA with repeated measures. There was one between-subject variable: Age-group (younger or older). There were two within-subject variables: Display (same level or different level) and Condition (Man-Man [MM], Bar-Bar [BB], Line-Line [LL], Man-Bar [MB], Line-Bar [LB], and Man-Line [ML]). There was a limited number of conditions in the design of this experiment. There was not an exhaustive set of all conditions as there would have been too many. The results of this analysis are presented in Table 5A.3.2.

The significant main effect of Display ($F(1,41)=75.56$, $p<0.001$) indicated that the same level display was easier than the different level display.

There was a significant main effect of Condition ($F(5,205)=6.28$, $p<0.001$). However, a Newman-Keuls Multiple Range test applied to this main effect demonstrated that each condition did not differ significantly from each other. A graphical representation of this main effect is presented in Figure 5A.3.1.

The significant interaction between Display and Condition ($F(5,205)=6.19$, $p<0.001$) showed that the effect of display affects significantly the children's performance on condition type. This interaction was explored using a Newman-Keuls Multiple Range Test (Bruning and Kintz,1977:119-

122). It was found that there was no difference in performance in all material conditions, except in the Man-Man Condition when they were presented in same level displays. For different level displays, there was a difference in performance in all conditions. Different material conditions were more difficult than same material conditions. Those conditions, for the different level display involving lines and bars, particularly those involving lines, resulted in worse performance than those involving only men. There was no difference in the Man-Man condition in either the same or different level displays. A graphical representation of this comparison is presented in Figure 5A.3.2.

The significant interaction between Condition and Age ($F(5,205)=2.64$, $p<0.05$) indicated that there was no difference between both age groups in the same type material condition. For the different type material condition, there was a difference in performance between both age groups. Younger children found different material conditions, particularly the Line-Bar condition, more difficult than did the older children. Again, the displays involving lines resulted in worst performance than the others, involving only men. This interaction was explored using a Newman-Keuls Multiple Range Test (Bruning and Kintz,1977:119-122). Graphical representations of this comparison is presented in Figure 5A.3.3.

5A.4 Discussion

The results of Experiment 5A indicated that children performed significantly better under the same level display than under the different level one.

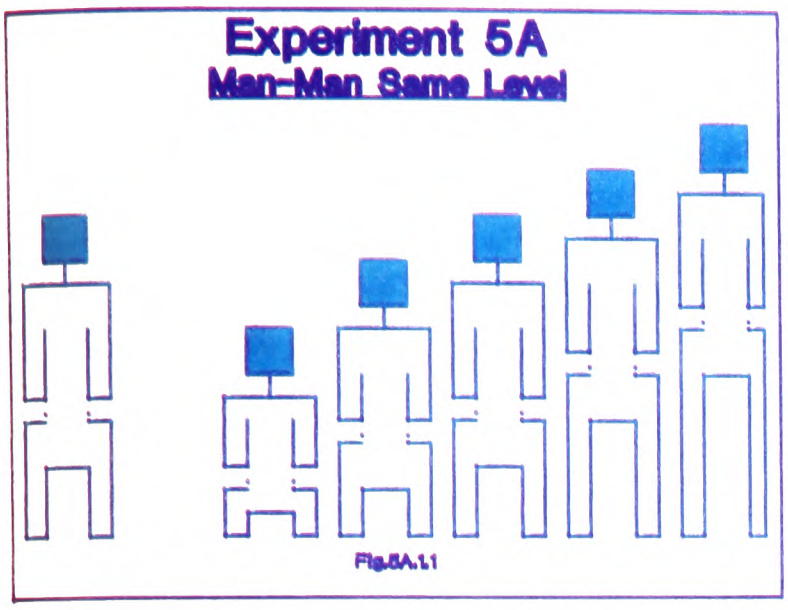
The same level display results in significantly better performance than does the different level display for all material conditions, except the Man-Man Condition. Hence, as predicted, children did not find it difficult to compare length at a different level in this type of MM display. However,

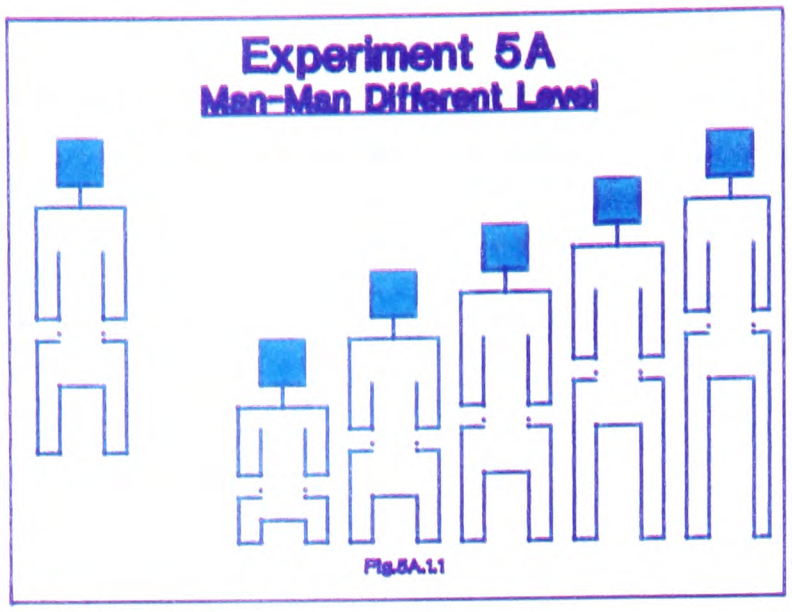
children found it difficult to compare length at a different level for all other display types.

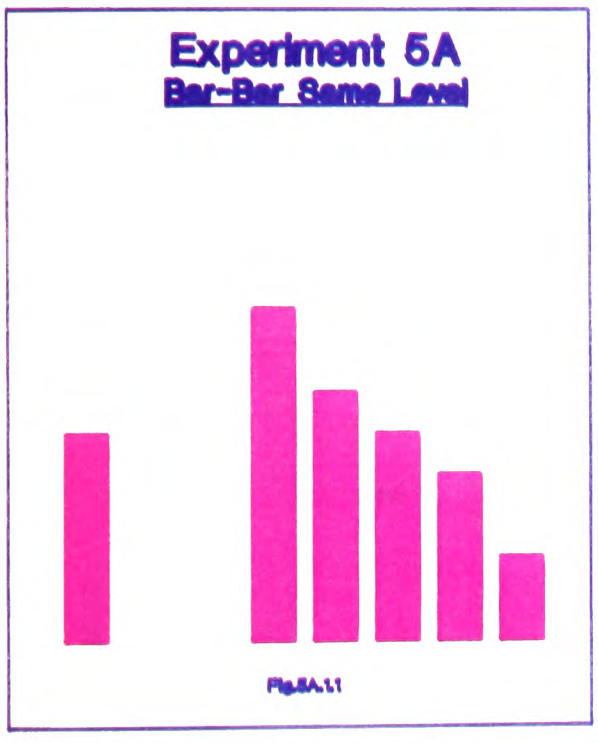
It seems that same level and different level displays have an effect on length comparison. Children seem to perform more successfully when comparing length if it is presented on a same level, than if it is presented on a different level.

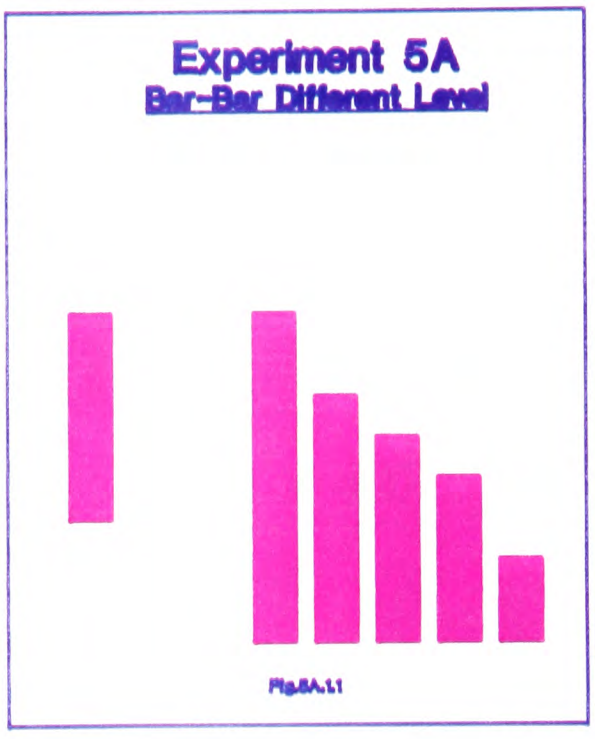
However, as there was no significant difference in the Man-Man (MM) Condition in either the same or different level displays, the second preliminary experiment (Experiment 5B) included only lines and bars (i.e. Line-Line (LL), Line-Bar (LB), Bar-Line (BL), and Bar-Bar (BB). Lines seem to have been affected by different level displays, more than men or bars. The next experiment (Experiment 5B) is an extension of the present one (Experiment 5A), but as mentioned previously, heterogeneity of material display will be curtailed.

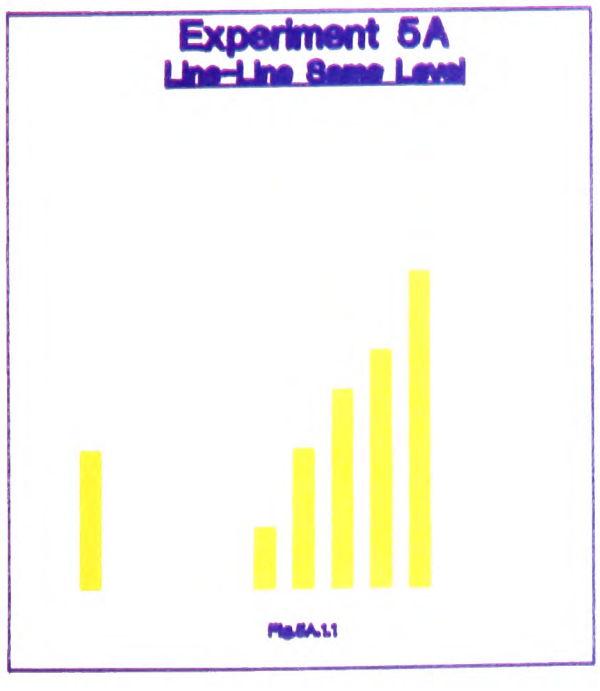
FIGURES AND TABLES FOR EXPERIMENT 5A

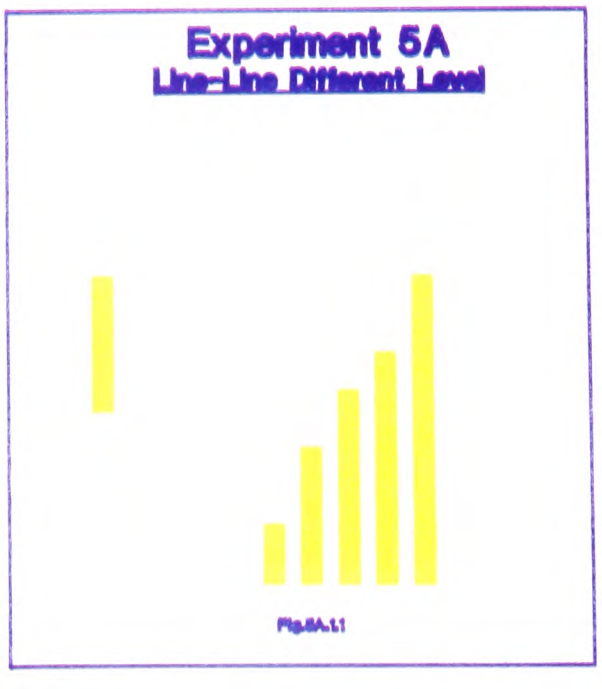

Experiment 5A
Man-Man Same Level
Fig.5A.1.1

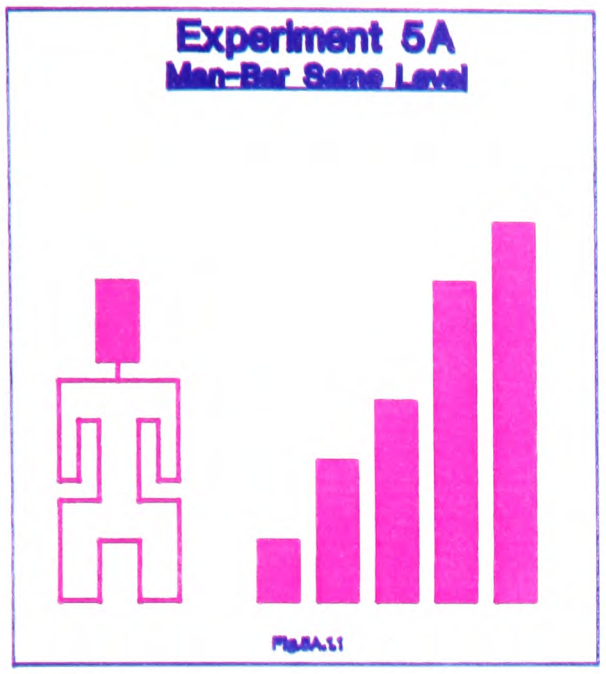

Experiment 5A
Man-Man Different Level
Fig.5A.1.1

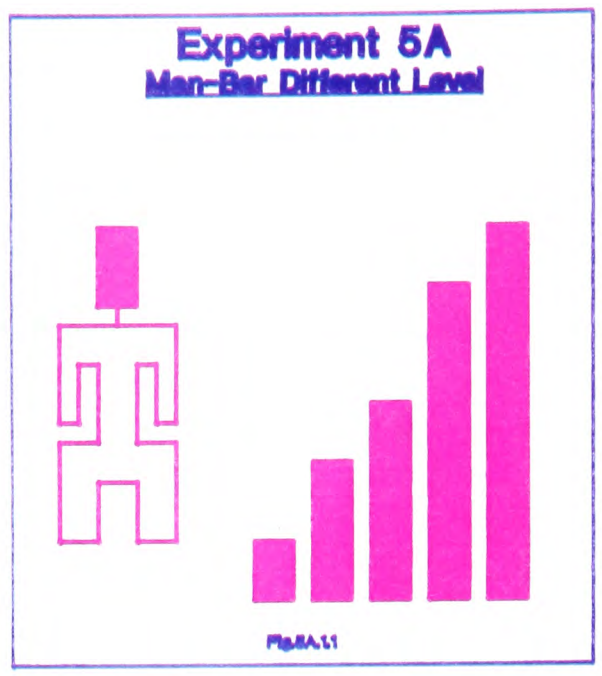

Experiment 5A
Bar-Bar Same Level
Fig.5A.1.1

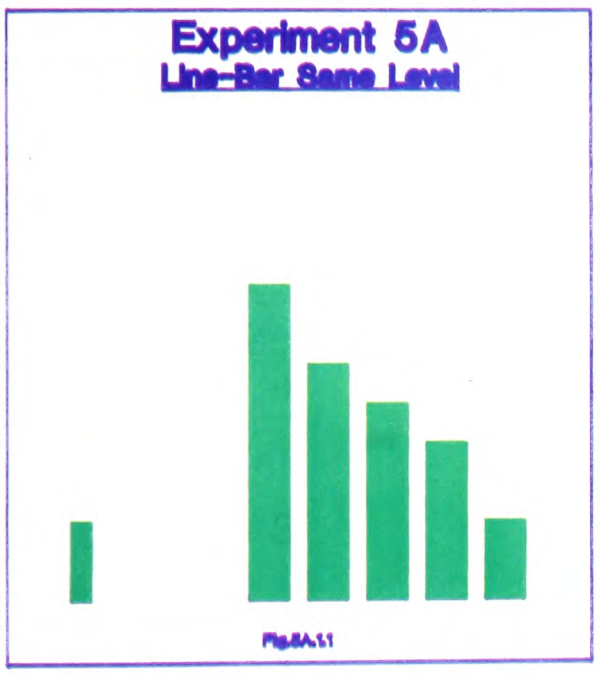

Experiment 5A
Bar-Bar Different Level
Fig.5A.1.1

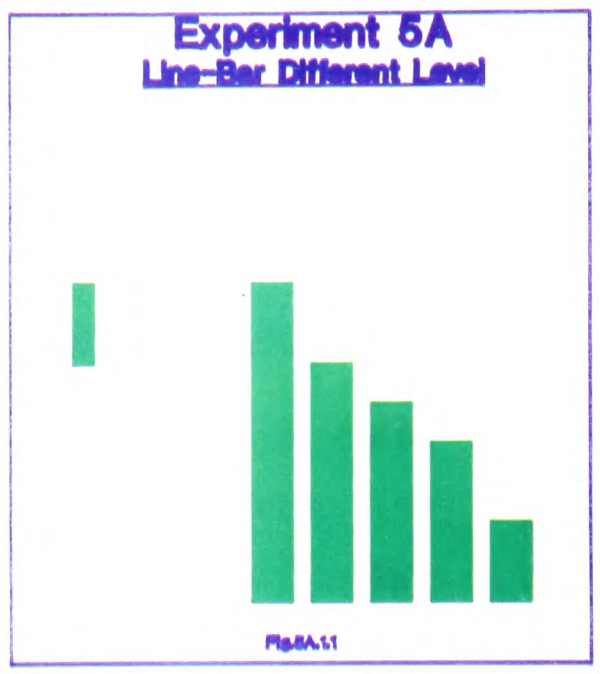

Experiment 5A
Line-Line Same Level
Fig.5A.1.1

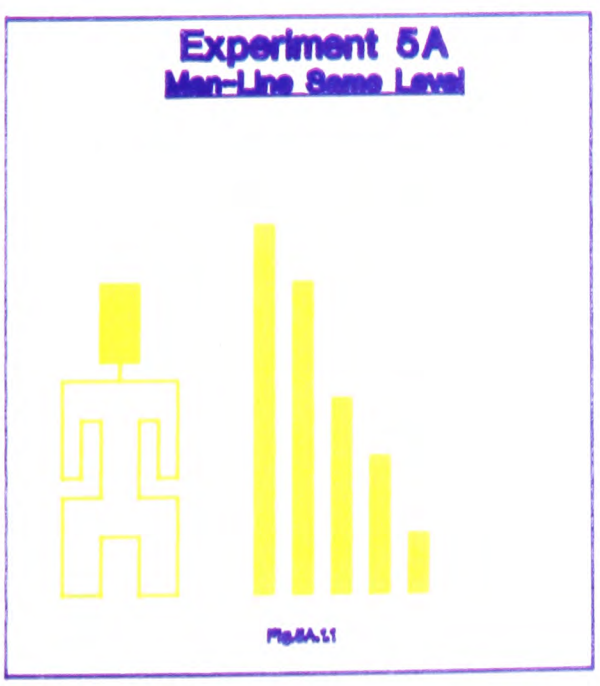

Experiment 5A
Line-Line Different Level
Fig.5A.1.1


Experiment 5A
Man-Bar Same Level
Fig.5A.1.1


Experiment 5A
Man-Bar Different Level
Fig.5A.1.1


Experiment 5A
Line-Bar Same Level
Fig.5A.1.1


Experiment 5A
Line-Bar Different Level
Fig.5A.1.1


Experiment 5A
Man-Line Same Level
Fig.5A.1.1


Experiment 5A
Man-Line Different Level
Fig.5A.1.1

Table 5A.2.1 The Design of Experiment 5A

No. of children	No.of children in each seq.	Material Sequence	Condition		Display (set order)
(Young group) 21	4	1	SAME	MM	A B
	4	2		BB	B A
	4	3		LL	A B
	3	4	DIFFERENT	MB	B A
	4	5		LB	A B
	2	6		ML	B A
(Old group) 22	4	1	SAME	MM	A B
	3	2		BB	B A
	4	3		LL	A B
	4	4	DIFFERENT	MB	B A
	3	5		LB	A B
	4	6		ML	B A

A = Set order Standing first, Raised second
B = Set order Raised first, Standing second

Table 5A.3.1 Means & Standard Deviations (out of 3) for the Number of Correct Responses

Age Group	Material Condition		Display Level Same	Display Level Different
Young	SAME	MM	2.71 (0.72)	2.43 (0.81)
		BB	2.52 (0.68)	2.10 (1.04)
		LL	2.71 (0.46)	2.00 (1.05)
	DIFFERENT	MB	2.76 (0.62)	2.19 (1.08)
		LB	2.43 (0.81)	1.57 (1.17)
		ML	2.62 (0.59)	1.48 (1.03)
Old	SAME	MM	2.68 (0.57)	2.27 (0.70)
		BB	2.50 (0.60)	2.18 (0.91)
		LL	2.82 (0.50)	2.23 (1.02)
	DIFFERENT	MB	2.64 (0.58)	2.27 (0.70)
		LB	2.73 (0.55)	2.36 (0.85)
		ML	2.59 (0.73)	1.36 (1.05)

(Standard deviations in brackets.)

Table 5A.3.2 Summary Table (ANOVA) for Total Correct Scores

Source	SS	DF	MS	F	PROB.
Mean	2823.899	1	2823.899	873.63	0.0000
AgeGroup[A]	1.108	1	1.108	0.34	0.5614
error	132.527	41	3.232		
Display[D]	47.357	1	47.357	75.56	0.0000
DA	0.474	1	0.474	0.76	0.3898
error	25.697	41	0.627		
Condition[C]	14.760	5	2.952	6.28	0.0000
CA	6.202	5	1.241	2.64	0.0246
error	96.395	205	0.470		
DC	10.264	5	2.053	6.19	0.0000
DCA	1.334	5	0.267	0.80	0.5480
error	68.030	205	0.332		

SS = Sum of Squared Deviations
DF = Degrees of Freedom
MS = Mean of Squared Deviations
F = Ratio of Variances
Prob. = Level of Significance

Experiment 5A
Mean Scores Condition

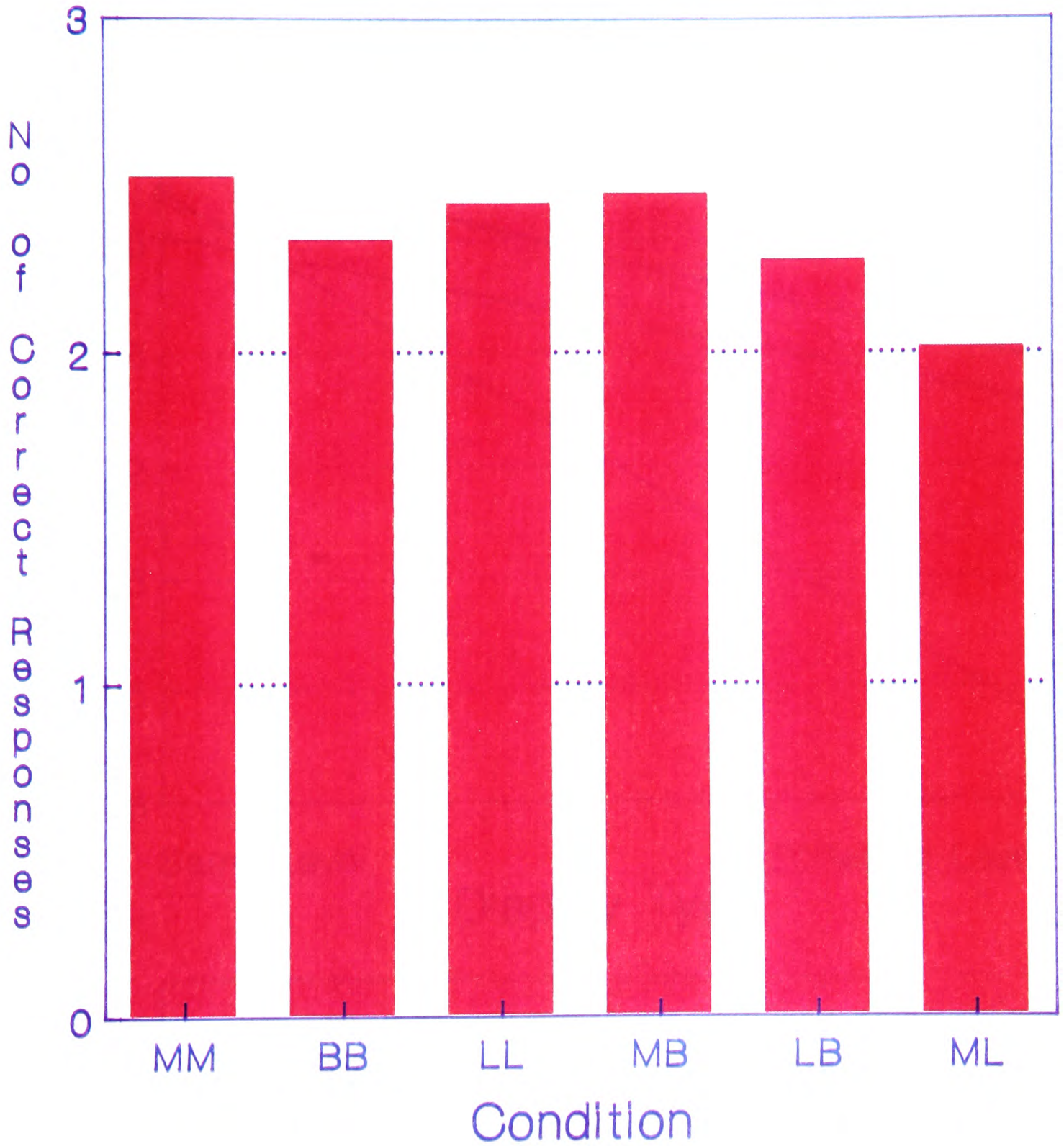


Fig.5A.3.1

Experiment 5A
Display x Condition

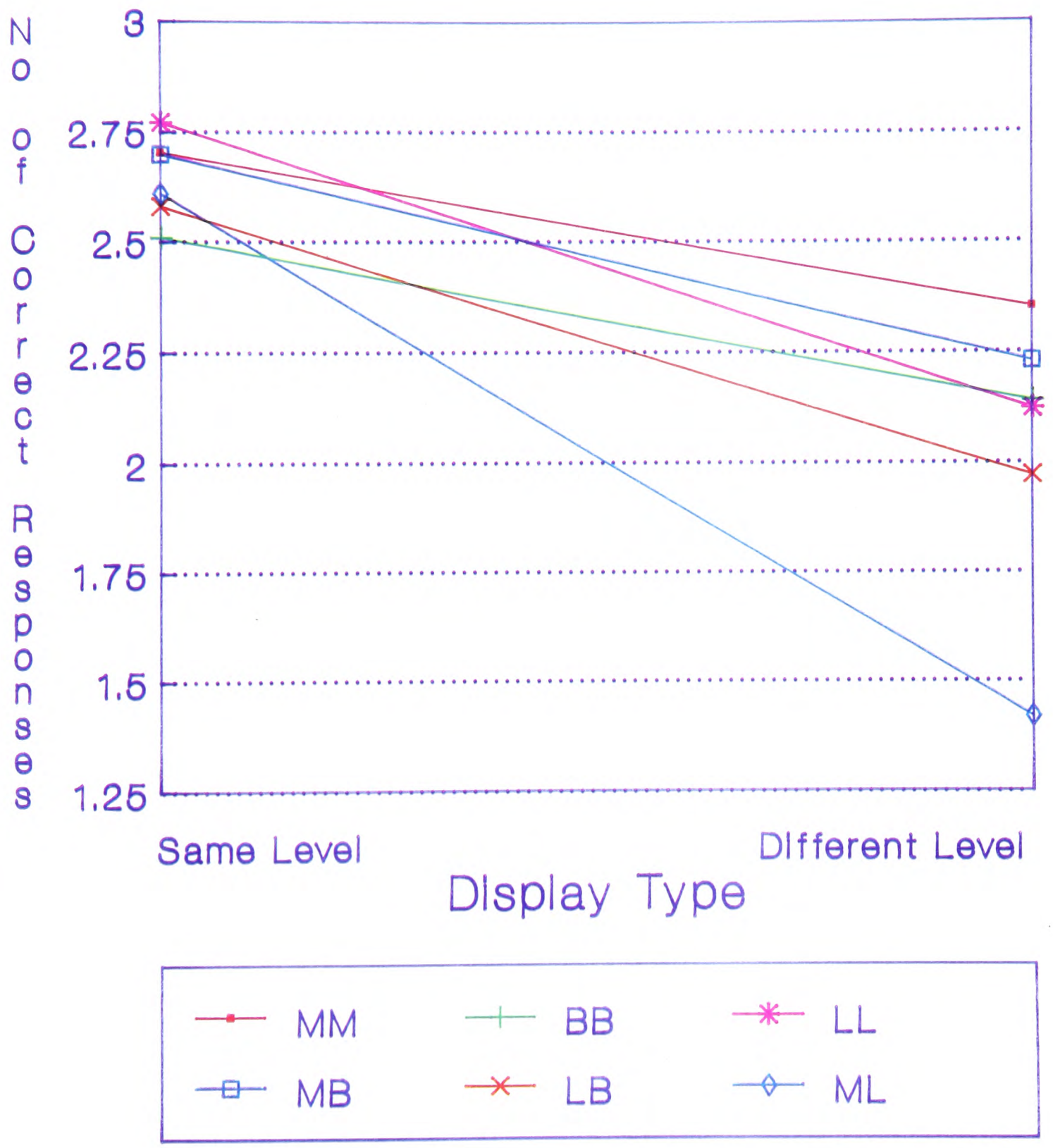


Fig.5A.3.2

Experiment 5A
Condition x Age

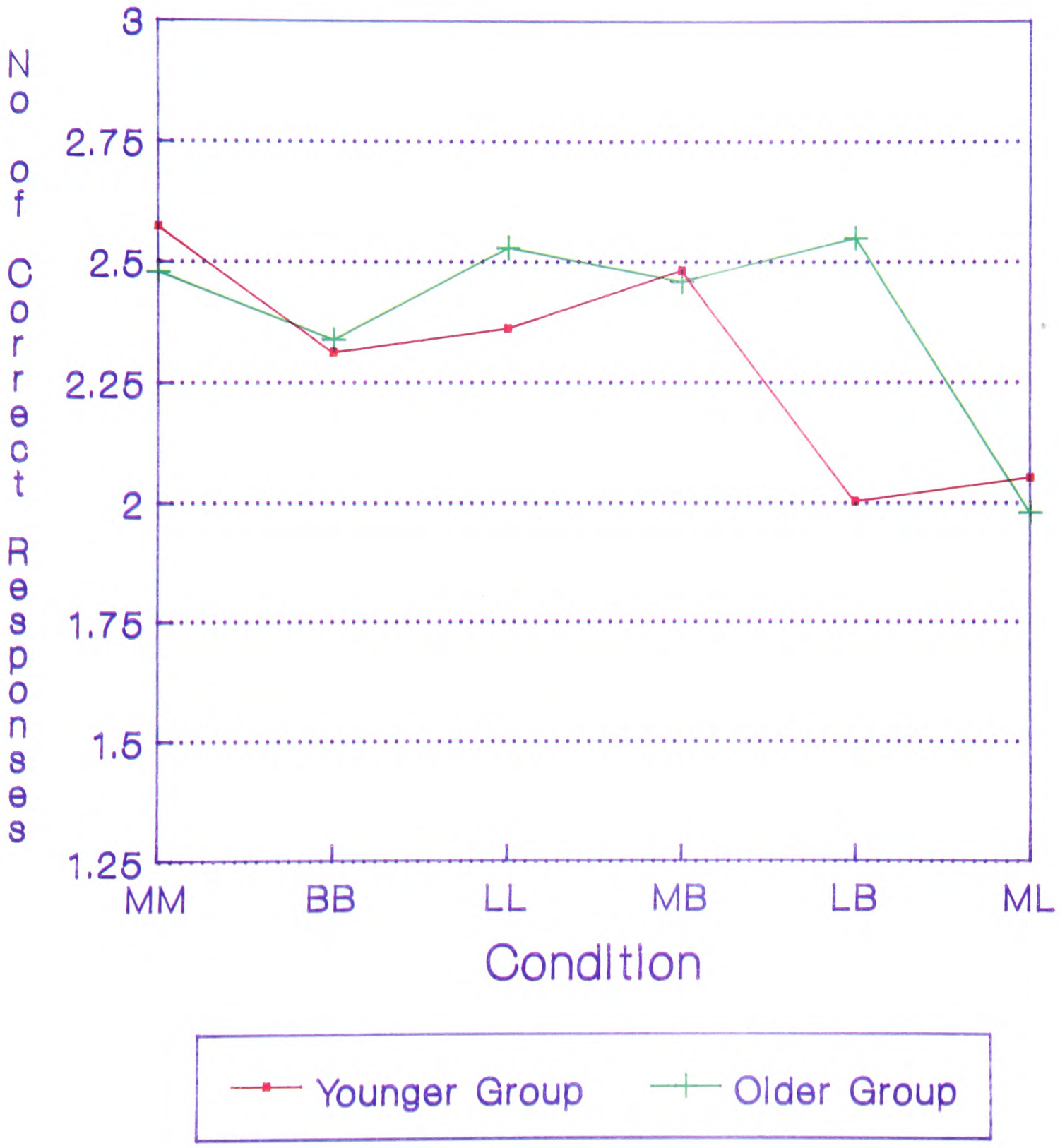


Fig.5A.3.3

APPENDIX FIVE

ANOTHER VIEW OF CHILDREN'S ABILITY TO REPRESENT ONE QUANTITY

APPENDIX FIVE

ANOTHER VIEW OF CHILDREN'S ABILITY TO REPRESENT ONE QUANTITY

Appendix 5 **Experiment 5B** **An Investigation of Responses made in Different Types of Displays using Additional Types of Continuous Material**

5B.1 Introduction

Experiment 5B is an extension of Experiment 5A. This is another preliminary experiment on length comparison. It continues to investigate children's success comparing length at different spatial levels. More specifically, it investigates children's performance in arriving at an absolute quantity via different spatial cues and stimulus heterogeneity.

The results of Experiment 5A proved that neither same nor different level displays had any significant effect on certain material type displays, such as Man-Man [MM]. Hence, this next experiment eliminates these displays and only deals with those displays (pertaining to the family of lines) which were significantly affected by the different level display.

A same level display, as in Experiment 5A, constituted the comparative and choice stimuli spatially arranged to begin at the same base level. A different level display, as in Experiment 5A, constituted the comparative and choice stimuli spatially arranged to begin at different base levels; in this spatial arrangement the comparative stimuli was higher than the choice stimuli. However, in contrast to Experiment 5A, it was the base level of the comparative stimuli that was raised to the height of the highest item of the choice stimuli. (See Figure 5B.1.1.) This spatial manipulation was devised to

confirm that children's success in the previous experiment, when arriving at an absolute quantity, was due to their ability in length comparison.

5B.2 Method

(a) Subjects

Two groups of children served as subjects: one consisting of 20 children ranging in age from 6;1 - 7;5 (mean age of 6;8), and the other of 21 children, ranging in age from 7;4 - 8;1 (mean age of 7;6). An additional group of 20 children (age-range 6;3 - 7;3; mean age 6;9) served as pilot subjects. The children were all attending a first school in Oxford, England.

(b) Apparatus

The same apparatus was used as for Experiments 2, 3, and 4.

(c) Design

Each child was given one sequence, in which there were 40 trials. Twenty of these 40 trials were for the same level display and the other 20 trials were for the different level display. As previously mentioned, a same level display, as in Experiment 5, constituted the comparative and choice stimuli spatially arranged to begin at the same base level. A different level display, as in Experiment 5, constituted the comparative and choice stimuli spatially arranged to begin at different base levels; in this spatial arrangement the comparative stimuli was higher than the choice stimuli. However, in contrast to Experiment 5, it was the base level of the comparative stimuli that was raised to the height of the highest item of the

choice stimuli. (See Figure 5B.1.1.) Each of the 40 trials was blocked into four different presentations. There were four displays. Each child was grouped into one of these orders. The four orders created for display type were the following:

Sequence 1	**Sequence 2**	**Sequence 3**	**Sequence 4**
BB	LL	BL	LB
LL	BL	LB	BB
BL	LB	BB	LL
LB	BB	LL	BL

where the material displays were:

(1) Condition BB. The comparative stimuli and the choice stimuli were bars. This was a within (same material) condition.

(2) Condition BL. The comparative stimuli was a bar and the choice stimuli were lines. This was a between (different material) condition.

(3) Condition LL. The comparative stimuli and the choice stimuli were lines. This was a within (same material) condition.

(4) Condition LB. The comparative stimuli was a line and the choice stimuli were bars. This was a between (different material) condition.

The display variable was presented in alternating order. For both groups of children, half of the subjects received the same level display first followed by the different level display; the other half were given the task in the reverse order. Hence, a mixed design with repeated measures on Conditions (3: Condition [C], Material [M], Display [D]) was used. Age (2: 6, 7) was the Between factor. The design for Experiment 5B is illustrated in Table 5B.2.1.

(d) **Materials**

The continuous computer-driven displays that were used are presented in Figure 5B.1.1. There were 10 displays (five in a same level display and an equivalent five in a different level display) for each of the four material conditions, described in the design section (c) of this experiment. Each display showed two sets of stimuli. Depending on the display, the comparative stimuli consisted either of (a) one bar or (b) one line. When a bar was presented, it was selected randomly from a set of 14 bars, which were 1,2,3,4,5,6,7,8,9,10,11,12,13, and 14 inches high. When a line was presented, it was selected randomly from a similar set of 14 lines. The choice stimuli consisted either of 5 bars or 5 lines. The comparative stimuli was always the same height as one of the choice stimuli.

In 10 displays, the choice stimuli were arranged in ascending order of height, while in the other 10 displays they were arranged in descending order of height. The choice stimuli that were arranged in ascending order in 10 of the displays were as follows:

2 sets of items that were 1, 3, 5, 7, and 9 inches in height
2 sets of items that were 2, 4, 6, 8, and 10 inches in height
2 sets of items that were 3, 5, 7, 9, and 11 inches in height
2 sets of items that were 4, 6, 8, 10, and 12 inches in height
2 sets of items that were 6, 8, 10, 12, and 14 inches in height

The choice stimuli that were arranged in descending order in 10 of the displays were as follows:

2 sets of items that were 10, 8, 6, 4, and 2 inches in height
2 sets of items that were 11, 9, 7, 5, and 3 inches in height
2 sets of items that were 12, 10, 8, 6, and 4 inches in height
2 sets of items that were 13, 11, 9, 7, and 5 inches in height
2 sets of items that were 14, 12, 10, 8, and 6 inches in height

Hence, in half of the displays the numbers of inches in height of the items were all odd, while in the other half they were all even.

In 20 displays per condition, the correct response was four times in each of the five positions of the choice stimuli. In all displays, the items (bars or lines) in each set were positioned one inch apart and there was a distance of two inches between the two sets. Any display in any condition could be green, pink, blue, or yellow in colour, and the colour of any given display was selected randomly.

Experiment 5B was different from Experiment 5A in the following points:

(1) The different level display varied. When the top level of the comparative stimuli was at the same level as the top of the tallest item of the choice stimuli, it constituted a different level display for Experiment 5A. When the base level of the comparative stimuli was at the same level as the top of the tallest item of the choice stimuli, it constituted a different level display for Experiment 5B.

(2) There were four conditions for Experiment 5B instead of six in Experiment 5A. These also included (1) a same material condition consisting of the same type of continuous displays as described in Experiment 5A (BB and LL) and in the design section (c) of this experiment, and (2) a different material condition consisting of a different type of continuous display as described in Experiment 5A (LB) and in the design section (c) of this experiment (LB and BL). (See Figure 5B.1.1.)

(e) Procedure

Children were taken from their classes individually, in a random order. The children were shown each display, one at a time as presented on the

screen. Each of the 20 displays was displayed twice under each of two display types (same level or different level).

Children were asked one question for each display. Hence, each child was asked 10 questions per material display (Condition BB, Condition LL, Condition BL, Condition LB), totalling 40 questions in all. The question asked was the same for each display.

The question took the form, *"Which one of these* (the experimenter pointed to the set of comparative stimuli) *is the same as this one* (pointing to the choice stimuli)*?"*. The child then selected an answer (one of the five bars or lines of the choice stimuli). Responses were recorded and stored in the computer.

5B.3 Results

The mean number of correct responses for each type of condition in each display type is shown in Table 5B.3.1. These means were obtained from scores on the five questions on each of the four types of material display and give an overall picture of children's performance. In general, more correct responses occurred in the same level display than in the different level display.

Table 5B.3.1 Means and Standard Deviations (out of 5) for the Number of Total Correct Responses

5B.3.1 Comparing Correct Scores across Conditions

The mean and standard deviations for the total number of correct responses made in each display for each material condition by the children were given in Table 5B.3.1. This table shows that both age groups elicited about the same amount of correct responses. Same level displays were superior to different level displays. Same type of material display yielded

more correct responses from the children than different type of material display. That is to say, children performed better in the BB and LL conditions than in the BL and LB conditions.

Calculating the Probability of Correct Responses occurring purely by Chance

As for Experiments 2, 3, and 4, children had a choice of five responses (from the choice stimuli). Therefore, there was a 20% chance of their choosing any one of these responses. Children were considered to be performing at chance level if they got a total of 1.0 correct response. The probability that the correct responses occurred simply by chance was explored using a modification of the standard hypothesis testing procedure with the means in Table 5B.3.1. (See Section 3.3.1.)

The proportions of correct responses made in all cell means were significantly ($p<0.01$) greater than expected purely by chance. This indicates that the children were not acting randomly.

5B.3.2 Analysis of Variance

These trends were analyzed by subjecting the raw scores for the total number of correct responses to a 2x2x2x2 ANOVA. There was one between-subject variable: age-group (younger or older). The other main terms were: Display (same level or different level), Material in Comparative Pair (Bars or Lines), and Condition (Comparative Pair and Choice Stimuli Same Material or Comparative Pair and Choice Stimuli Different Material), with repeated measures on these three factors. The results of this analysis are presented in Table 5B.3.2.

The significant main effect of Display ($F(1,39)=21.66$, $p<0.001$) indicated that the same level display was easier than the different level display.

The significant main effect of Condition ($F(1,39)=16.90$, $p<0.001$) indicated that the children performed better when same type displays were presented than when different type displays were presented. Children were better in the BB and LL conditions than the BL and LB conditions.

An interaction between Display and Condition ($F(1,39)=6.07$, $p<0.05$) appeared. This interaction showed that the effect of display affects significantly the children's performance on condition type. This interaction was explored using a Newman-Keuls Multiple Range Test (Bruning and Kintz,1977:119-122). It was found that performance on different material under the different level display was significantly worse than performance on different material under the same level display. Conditions BL and LB under the different level display had the worst performance. Same level and different level displays did not have a significant effect on same material [LL and BB]. A graphical representation of this comparison is presented in Figure 5B.3.1.

5B.4 Discussion

The results of Experiments 5B indicated that children performed significantly better under the same level display than under the different level display. Children also performed significantly better when the two sets of items were of the same material than when they were of different material. In particular, the interaction between Display and Condition demonstrated that children performed worse on different material under the different level display. The results replicated those of Experiment 5A.

These results also indicated that children habituate to the mentioned-tasks. This is evident in the lack of age difference among these tasks. There was no difference in performance between the younger group and the older group for Experiment 5B.

It is clear that children are confused by spatial cues and stimuli heterogeneity, but nevertheless can and do arrive, at the absolute value of a quantity, particularly when same material and same level displays are presented. Children perceive spatial arrangement and heterogeneity as relevant to numerical size, as demonstrated in their difficulty with different type material in the different level display. This seems to affect them in the development of concepts of numerical equivalence.

One cannot help but wonder what the results would be like in an experiment of numerical comparison rather than numerical equivalence. Experiment 6 was then designed to investigate whether these spatial cues and stimuli heterogeneity are crucial factors in the development of concepts of quantitative comparison. Of particular concern were the Equalize and Compare questions.

The enigma of what is so difficult about comparing different types of material and what makes this comparison particularly difficult when the material is on a different level, still remains.

Two possible hypotheses may explain this enigma. One hypothesis is attributed to "Action". A child thinks in terms of action. Thus, the two different materials cannot be satisfactorily compared. One cannot grow out of another; the changes cannot be seen as normal growth, as one item can not be placed directly on top of another.

A second hypothesis is attributed to "Appearance". Moore and Frye (1986) [see Section 1.7.3] suggested the possibility that children construe a task differently in two contexts. In a standard context (in these experiments: same level and same material), the task is understood to require a

simultaneous comparison of the two amounts on the basis of their appearance. Thus, a change in appearance (i.e. in these experiments, different level and different material) is taken to mean a change in amount. Therefore, this implies that young children use information about the appearance when making judgements of the display. The overall conclusion is that a child's understanding of context is inseparable from the understanding of quantity itself.

These hypotheses are more relevant to quantitative comparison than to absolute quantity. Experiment 6 is aimed at testing the validity of these and related hypotheses, as well as investigating how they apply to quantitative comparison.

FIGURES AND TABLES FOR EXPERIMENT 5B

Experiment 5B
Bar-Bar Same Level

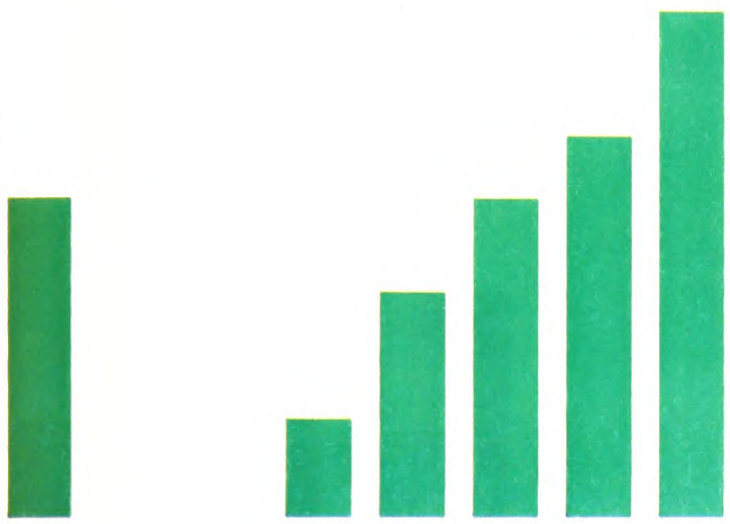


Fig.5B.1.1

Experiment 5B
Bar-Bar Different Level

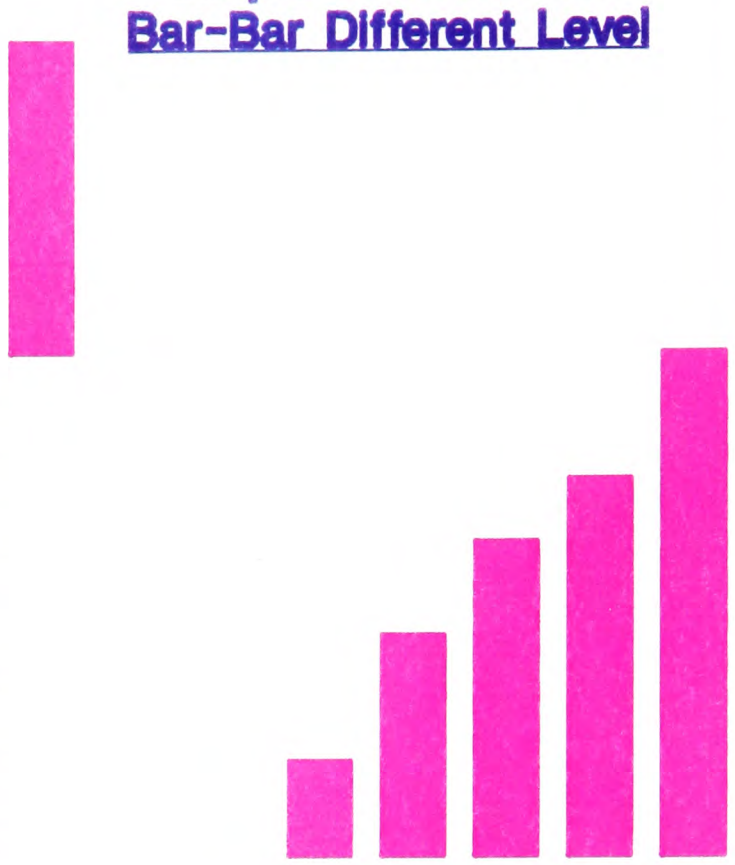


Fig.5B.1.1

Experiment 5B
Line-Line Same Level

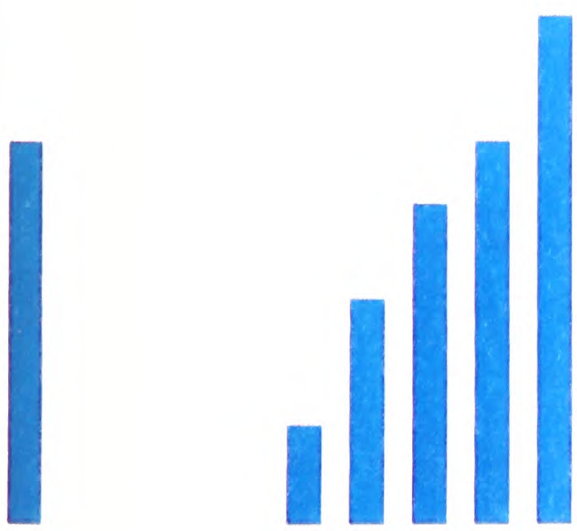


Fig.5B.1.1

Experiment 5B
Line-Line Different Level

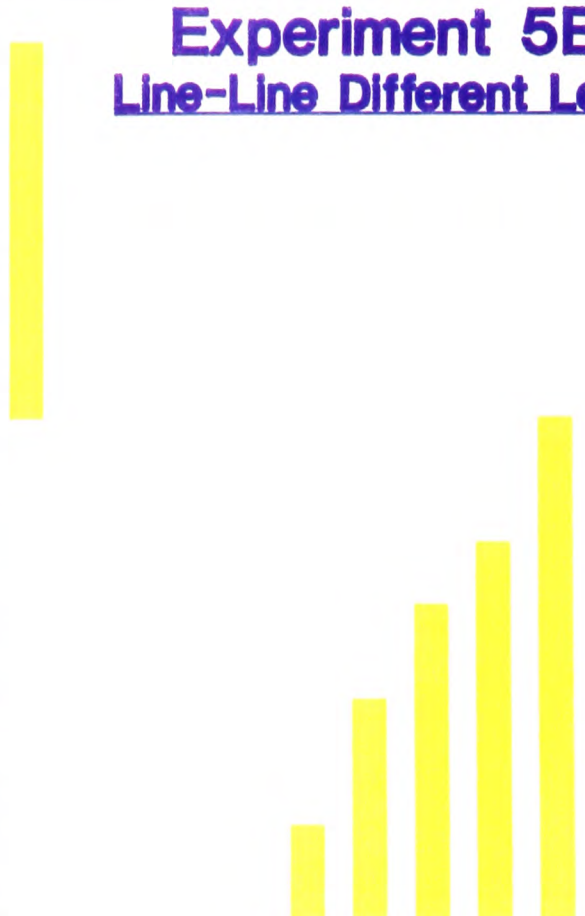


Fig.5B.1.1

Experiment 5B
Bar-Line Same Level

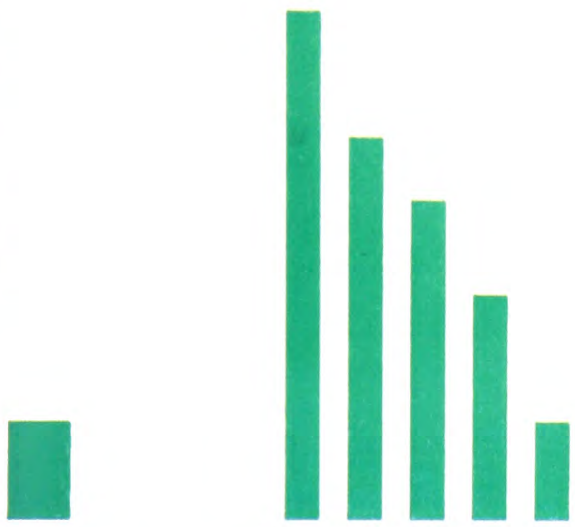


Fig.5B.1.1

Experiment 5B
Bar-Line Different Level

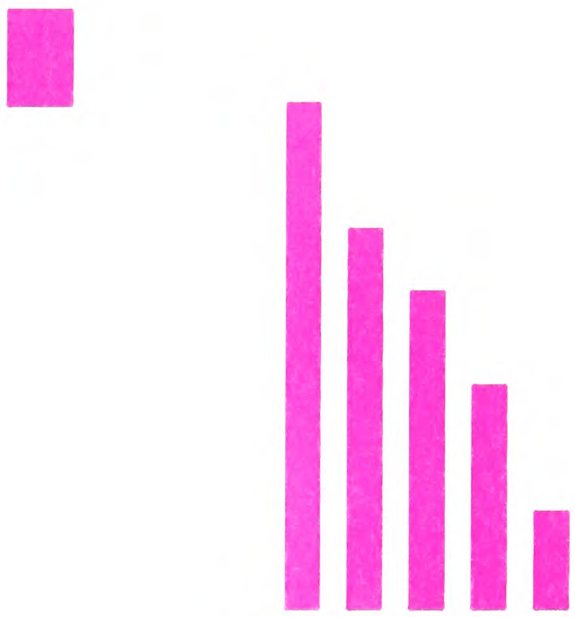


Fig.5B.1.1

Experiment 5B
Line-Bar Same Level

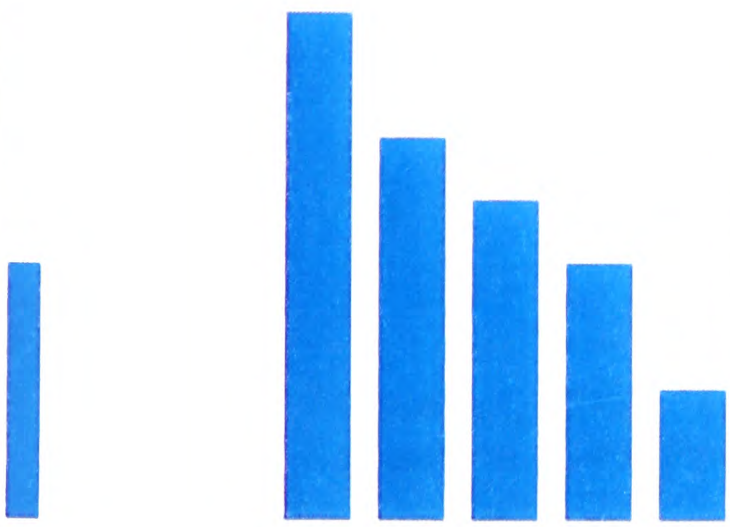


Fig.5B.1.1

Experiment 5B
Line-Bar Different Level

Fig.5B.1.1

Table 5B.2.1 The Design of Experiment 5B

No. of children	No. of children in each seq.	Material Sequence	Display		Condition (Seq.Set)
	6	1	SAME	BB	A B
	5	2		LL	B A
(Young Group) 20	5	3	DIFFERENT	BL	A B
	4	4		LB	B A
	5	1	SAME	BB	A B
	5	2		LL	B A
(Old Group) 22	6	3	DIFFERENT	BL	A B
	5	4		LB	B A

A = Sequence Set Standing first, Raised second
B = Sequence Set Raised first, Standing second

Table 5B.3.1 Means & Standard Deviations (out of 5) for the Number of Correct Responses

Age Group	Material Group		Display: Same	Display: Different
Young	SAME	BB	4.95 (0.22)	4.75 (0.55)
		LL	4.85 (0.37)	4.75 (0.44)
	DIFFERENT	BL	4.60 (1.00)	3.85 (1.42)
		LB	4.65 (0.67)	4.10 (0.97)
Old	SAME	BB	4.76 (0.54)	4.62 (0.97)
		LL	4.86 (0.66)	4.67 (0.73)
	DIFFERENT	BL	4.57 (1.03)	4.29 (0.85)
		LB	4.76 (0.89)	4.52 (0.93)

(Standard deviations in brackets.)

Table 5B.3.2 Summary Table (ANOVA) for Total Correct Scores

Source	SS	DF	MS	F	PROB.
Mean	6926.482	1	6929.482	2191.85	0.0000
AgeGroup[A]	0.384	1	0.384	0.12	0.7293
error	123.244	39	3.160		
Display[D]	7.731	1	7.731	21.66	0.0000
DA	0.707	1	0.707	1.98	0.1674
error	13.921	39	0.357		
Condition[C]	10.488	1	10.488	16.90	0.0002
CA	2.293	1	2.293	3.69	0.0619
error	24.201	39	0.621		
DC	1.815	1	1.815	6.07	0.0182
DCA	0.839	1	0.839	2.81	0.1018
error	11.655	39	0.299		
Material[M]	0.762	1	0.762	2.42	0.1281
MA	0.177	1	0.177	0.56	0.4586
error	12.293	39	0.315		
DM	0.115	1	0.115	0.43	0.5173
DMA	0.115	1	0.115	0.43	0.5173
error	10.525	39	0.270		
CM	0.602	1	0.602	3.08	0.0874
CMA	0.017	1	0.017	0.09	0.7716
error	7.636	39	0.196		
DCM	0.049	1	0.049	0.33	0.5667
DCMA	0.000	1	0.000	0.00	0.9888
error	5.701	39	0.146		

SS = Sum of Squared Deviations
DF = Degrees of Freedom
MS = Mean of Squared Deviations
F = Ratio of Variances
Prob. = Level of Significance

Experiment 5B
Display x Condition

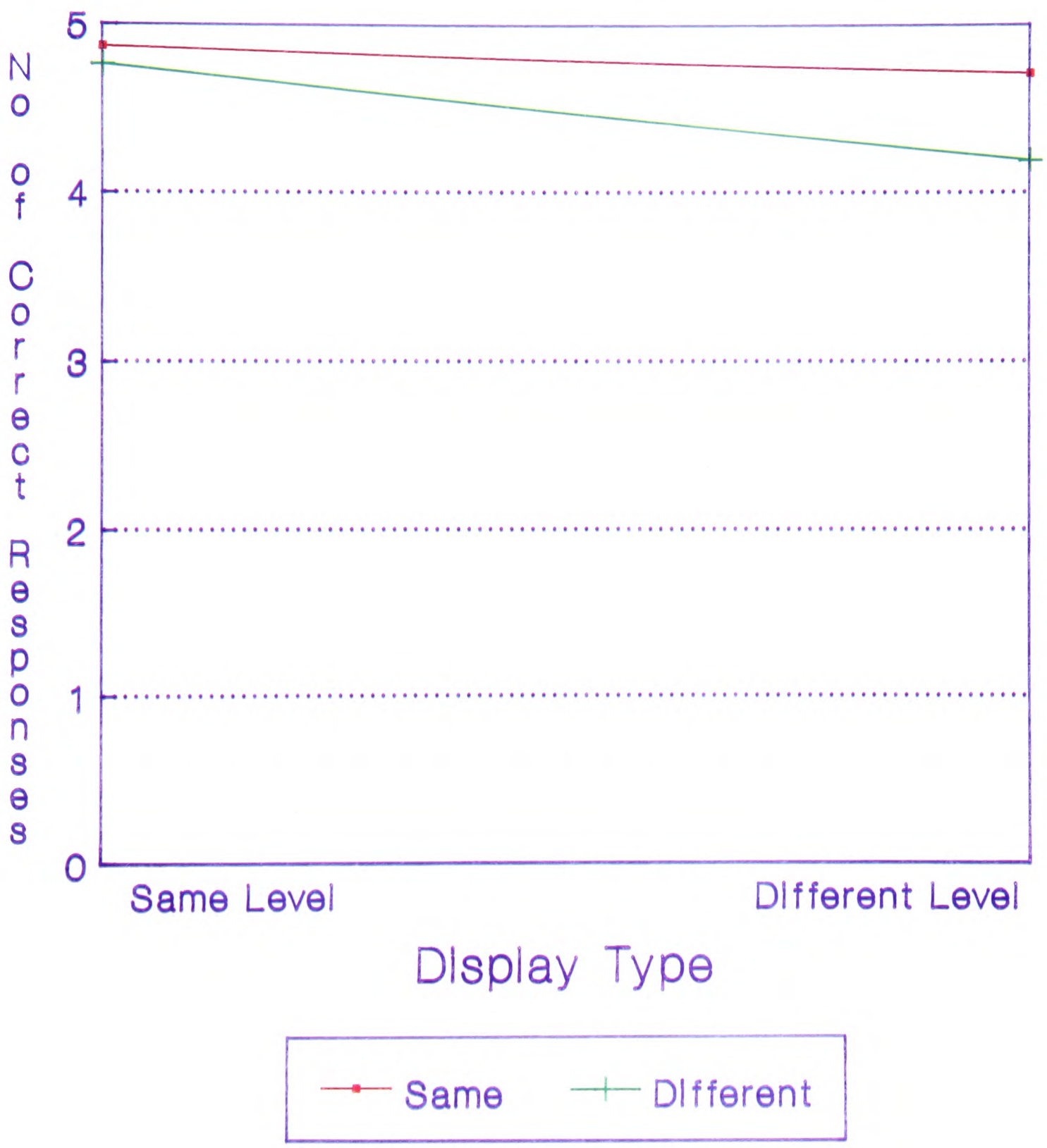


Fig.5B.3.1

APPENDIX SIX

STATISTICAL TABLES FOR EXPERIMENT 6

I. Means and Standard Deviations (out of 3) for the Number of Total Correct Responses for Experiment 6

II. Summary Table (ANOVA) for Total Correct Scores for Experiment 6

Table 7.3.1 Means and Standard Deviations (out of 3) for the Number of Correct Responses

Age Group	Display	Comparative Term							
		More				Less			
		Equalize		Compare		Equalize		Compare	
		Same	Diff	Same	Diff	Same	Diff	Same	Dif
Young	BB	1.52	1.67	1.67	1.52	1.62	1.86	1.57	1.48
		1.03	*0.86*	*0.97*	*0.98*	*0.87*	*0.91*	*0.81*	*0.98*
	LL	1.52	1.57	1.67	1.71	1.29	1.48	1.24	1.67
		0.93	*0.98*	*0.86*	*0.90*	*0.78*	*0.81*	*0.94*	*0.97*
	BL	1.86	1.71	1.52	1.67	1.52	1.57	1.52	1.81
		1.06	*0.90*	*0.75*	*1.07*	*0.93*	*1.08*	*0.98*	*0.93*
	LB	1.67	1.81	1.48	1.43	1.57	1.38	1.29	1.43
		0.97	*0.98*	*0.93*	*0.68*	*1.03*	*1.11*	*0.90*	*0.87*
Old	BB	2.53	2.47	2.21	2.58	2.47	2.26	2.32	2.53
		0.77	*0.77*	*0.86*	*0.77*	*0.77*	*1.05*	*0.89*	*0.70*
	LL	2.32	2.47	2.58	2.32	2.16	2.32	2.21	2.42
		0.75	*0.70*	*0.69*	*0.82*	*1.02*	*0.82*	*0.79*	*0.84*
	BL	2.47	2.26	2.21	2.53	2.16	2.21	2.37	2.37
		0.70	*0.81*	*0.79*	*0.70*	*0.96*	*0.86*	*0.76*	*0.83*
	LB	2.21	2.32	2.32	2.16	2.16	2.42	2.26	2.32
		0.79	*0.86*	*0.82*	*0.96*	*0.96*	*0.77*	*0.93*	*0.82*

(Standard deviations in italics.)

Table 7.3.4 Summary Table (ANOVA) for Total Correct Scores

Source	SS	DF	MS	F	PROB.
Mean	4884.678	1	4884.678	383.80	0.0000
AgeGroup[A]	188.778	1	188.778	14.83	0.0004
error	483.627	38	12.727		
ComparativeTerm[Ct]	2.297	1	2.297	1.16	0.2883
CtA	0.159	1	0.159	0.08	0.7784
error	75.259	38	1.981		
Question[Q]	0.071	1	0.071	0.07	0.7938
QA	0.634	1	0.634	0.61	0.4378
error	39.178	38	1.031		
CtQ	0.427	1	0.427	0.34	0.5652
CtQA	0.052	1	0.052	0.04	0.8406
error	48.191	38	1.268		
Display[D]	1.561	1	1.561	3.91	0.0553
DA	0.018	1	0.018	0.04	0.8345
error	15.182	38	0.400		
CtD	0.552	1	0.552	2.74	0.1063
CtDA	0.046	1	0.046	0.23	0.6363
error	7.666	38	0.202		
QD	0.180	1	0.180	0.51	0.4807
QDA	0.011	1	0.011	0.03	0.8610
error	13.470	38	0.355		
CtQD	0.118	1	0.118	0.39	0.5345
CtQDA	0.161	1	0.161	0.54	0.4672
error	11.363	38	0.299		
Condition[C]	0.472	1	0.472	2.37	0.1317
CA	0.810	1	0.810	4.07	0.0508
error	7.558	38	0.199		
CtC	0.012	1	0.012	0.03	0.8686
CtCA	0.287	1	0.287	0.65	0.4244
error	16.737	38	0.441		

Table 7.3.4
Continuation Summary Table (ANOVA)

Source	SS	DF	MS	F	PROB.
CtQC	0.775	1	0.775	3.14	0.0845
CtQCA	0.675	1	0.675	2.73	0.1065
error	9.381	38	0.247		
DC	0.125	1	0.125	0.48	0.4915
DCA	0.032	1	0.032	0.12	0.7292
error	9.868	38	0.260		
CtDC	0.032	1	0.032	0.14	0.7150
CtDCA	0.125	1	0.125	0.54	0.4681
error	8.868	38	0.233		
QDC	0.103	1	0.103	0.46	0.5035
QDCA	0.721	1	0.721	3.21	0.0814
error	8.553	38	0.225		
CtQDC	0.159	1	0.159	0.59	0.4475
CtQDCA	0.515	1	0.515	1.91	0.1752
error	10.259	38	0.270		
Material[M]	2.630	1	2.630	6.75	0.0133
MA	0.255	1	0.255	0.65	0.4234
error	14.807	38	0.390		
CtM	0.431	1	0.431	1.27	0.2662
CtMA	0.693	1	0.693	2.05	0.1604
error	12.850	38	0.338		
QM	0.006	1	0.006	0.01	0.9059
QMA	0.018	1	0.018	0.04	0.8348
error	15.881	38	0.418		
CtQM	0.105	1	0.105	0.40	0.5315
CtQMA	0.105	1	0.105	0.40	0.5315
error	9.957	38	0.262		
DM	0.036	1	0.036	0.15	0.6997
DMA	0.017	1	0.017	0.07	0.7903
error	8.976	38	0.236		

Table 7.3.4
Continuation Summary Table (ANOVA)

Source	SS	DF	MS	F	PROB.
CtDM	0.388	1	0.388	1.08	0.3055
CtDMA	0.532	1	0.532	1.48	0.2314
error	13.655	38	0.359		
QDM	0.882	1	0.882	4.06	0.0509
QDMA	2.163	1	2.163	9.97	0.0031
error	8.243	38	0.217		
CtQDM	1.200	1	1.200	3.66	0.0631
CtQDMA	0.019	1	0.019	0.06	0.8101
error	12.449	38	0.328		
CM	0.016	1	0.016	0.04	0.8409
CMA	0.091	1	0.091	0.24	0.6281
error	14.421	38	0.380		
CtCM	0.670	1	0.670	2.30	0.1380
CtCMA	0.033	1	0.033	0.11	0.7401
error	11.092	38	0.292		
QCM	1.668	1	1.668	5.89	0.0201
QCMA	0.168	1	0.168	0.59	0.4458
error	10.762	38	0.283		
CtQCM	0.121	1	0.121	0.98	0.3276
CtQCMA	0.084	1	0.084	0.68	0.4148
error	4.690	38	0.123		
DCM	0.152	1	0.152	0.45	0.5050
DCMA	0.321	1	0.321	0.95	0.3347
error	12.784	38	0.336		
CtDCM	0.468	1	0.468	2.31	0.1372
CtDCMA	0.075	1	0.075	0.37	0.5485
error	7.719	38	0.203		
QDCM	0.296	1	0.296	1.64	0.2079
QDCMA	0.703	1	0.703	3.89	0.0558
error	6.859	38	0.181		

Table 7.3.4
Continuation Summary Table (ANOVA)

Source	SS	DF	MS	F	PROB.
CtQDCM	0.061	1	0.061	0.22	0.6455
CtQDCMA	0.005	1	0.005	0.02	0.8944
error	10.838	38	0.285		

SS = Sum of Squared Deviations
DF = Degrees of Freedom
MS = Mean of Squared Deviations
F = Ratio of Variances
Prob. = Level of Significance