EFFECTS OF DIFFERENTIAL ARITHMETIC PRACTICE ON CHILDREN'S SOLUTIONS OF MATHEMATICAL WORD PROBLEMS

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Marie Christine Rodriguez

August 2004

EFFECTS OF DIFFERENTIAL ARITHMETIC PRACTICE ON CHILDREN'S SOLUTIONS OF MATHEMATICAL WORD PROBLEMS

Marie Christine Rodriguez, Ph.D.

Cornell University 2004

Word problems are difficult. Although children eventually master computational skills, problem solving skills remain poor through adulthood. Two different types of manipulations were attempted to affect rates of successful word problem solution. First we made changes to the word problems themselves to make them more comprehensible for students, and therefore easier to solve. Second, students were given one of two types of arithmetic practice and were compared with a third group of students who received no additional practice to determine whether such practice could assist students with solving arithmetic word problems.

First- and second-grade students were tested on three different types of single-step arithmetic word problems: a set of Compare problems, a set of six typically worded Change problems and a set of six Change problems whose wording was clarified with simple temporal, semantic and referential clarifications. These changes were intended to make the action in the problem easier to follow so students could model the problems more successfully. The percentage of students answering correctly on different problems was compared.

Students were then randomly assigned to one of two different arithmetic worksheet conditions or to a third no practice condition. Worksheets consisted of either standard arithmetic practice or computational

practice requiring students to solve for something other than the result. After completing all of the worksheets, students were tested on a set of word problems arithmetically identical to those presented five months earlier.

Results of clarification were mixed. Students had somewhat more difficulty with solve-for-result problems which are traditionally the type of word problems at which students perform best. Students were more successful at solving clarified solve-for-start-set problems. There was also a curious trend for students to be more successful at subtraction problems than addition problems of the same type. This was more pronounced with clarified problems.

Second-grade students showed no effect of worksheet condition. First-grade students who were assigned to the non-canonical worksheet condition demonstrated a marked improvement on typically worded change problems. Reasons why the arithmetic practice did not also have an effect on clarified problems need to be explored further.

BIOGRAPHICAL SKETCH

Marie Christine Rodriguez was born on the eighth of March 1965 in Takoma Park, Maryland, the eldest of two daughters. She was named for Marie Curie and has had a lifelong fascination with science and to a lesser extent, mathematics. Her family moved to California in 1968 where she grew up in San Diego. She decided when she was three years old that she was going to become a doctor and did not waiver from that decision until late in college. She was a member of the third graduating class of Samuel Gompers Secondary School, a science, mathematics and computer magnet school that drew an eclectic mix of some of the city’s best and brightest students. Marie graduated from high school with honors and returned to the East Coast to attend college where she received her Bachelors degree in Natural Sciences from The Johns Hopkins University in 1987, concentrating in Biology and Physics.

After graduating from college, she did a number of things before returning to graduate school including working as a technical writer. She also spent a year in Kumamoto, Japan teaching English to middle school students with the JET Program. After returning to the United States, Marie lived in the Boston area for several years where she spent time working at the American Academy of Arts and Sciences, the oldest honorary society in the United States, and for one of the numerous of start-up computer firms of the early 90s. She eventually applied for graduate school, and opted to attend Cornell University. She received her M.A. in Human Experimental Psychology in 1998.

For my parents, John and Mihoko Rodriguez, who brought us up to believe
that nothing is more important than education,
and for Bonnie Dalzell, M.A., who demanded that I not join the ranks of those
whose graduate degrees were lost to the pursuit of purebred dogs and thus
repeat her greatest regret.

ACKNOWLEDGMENTS

Tim DeVoogd observed when I was interviewing at Cornell for graduate school that there was generally a higher correlation between GPA and GRE scores than there was in my case. This observation was unfortunately accurate and no small part of the reason that actually going back to graduate school was an almost unrealized dream. My undergraduate transcript is somewhat colorful, since study habits had been unnecessary for most of my life until then and I had never developed any though I did manage to graduate in four years. Like many people, I spoke of going back to graduate school eventually, but year after year passed without my developing a firm idea of what I wanted to study beyond that I was interested generally in how children think and learn.

Professor Ellen Winner helped focus my 'I'd like to go back to graduate school…someday' into reality. I had applied to the Psychology Department at Boston College on a lark while applying for graduate school in science education mostly because the department actively encouraged applicants who had world experience and did not care if they had a background or an undergraduate degree in psychology. Although I did not get in that year, instead of sending a rejection letter, Professor Winner very kindly called to explain that they had had three slots open and although I was the first person on their waiting list, all three students had accepted their offers. She also offered suggestions for how I could strengthen my application for the following year if I chose to reapply. I took her advice, and her Advanced Topics in Development course from which I discovered that I was particularly interested understanding how about scientific, numerical and mathematical reasoning in children, especially young children. I applied to several psychology programs the following year and chose to attend Cornell.

The Keil-Spelke lab group and its subsequent incarnations provided lively and challenging discussions in broad ranging topics in cognition and development. Without these weekly discussions I would not have nearly the breadth in cognition that I have.

I am particularly indebted to Grant Gutheil who was a post-doctoral fellow in the department during my first two years of graduate school. I am convinced that his advice and friendship – not to mention his usually open door and occupied office – kept me from giving up on graduate school when I became frustrated and/or discouraged. Grant has a unique capacity for being able to condense what he has heard into a concise and organized summary after listening to someone ramble on for 10 or 20 minutes. These visits greatly assisted someone with far too little writing experience to learn skills I should already have had. Grant also introduced me to his Adventure friends, a group of people scattered mostly across upstate New York who, in spite of the fact I see many of them no more than a single weekend a year, have become some of my closest friends as well. In particular, Nick and Sue Hogan always had an extra place at the table, lively and thoughtful conversation and a spare bed if it grew late. Their three children, Maeve, Conor and Brigid, always had warm hugs and enthusiasm to spare, and there was always at least one of them in the mood for whatever sort of company I was at the moment.

Alicia Cool, M.D. introduced me to the principal of the school at which these data were collected. I am enormously grateful to that principal for making access to her school and students simple, to the teachers who allowed me into their classrooms, to the children who participated in the “extra math”, and to the parents who allowed their children to participate. Mark Komisky's gold couch was my home-during-the-week on and off for much of the academic year during which these data were collected and his friendship which has now spanned half of my life has lent me fortitude through many things, this dissertation process not the least of them.

Assorted undergraduates in the Cornell Infant Research Center assisted with indexing data tapes and basic data coding.

They say that drug pushers give you the first taste for free…although she insists that Lilly, being broken and needing hours of patient rehabilitation, was not precisely free, Bonnie Dalzell did in fact not charge me a cent for my first borzoi. Bonnie has provided many hours of spirited discussion on a wide range of topics, many though not all related in some way to purebred dogs and to borzoi in particular. When my home in Ithaca was sold, she offered me a place to live and write, knowing from personal experience how easy it is to not quite finish a dissertation. She has been an extraordinary friend and mentor in addition to being the breeder of my dogs, who have themselves provided sanity, humor and unconditional love (if in fur suits). When I made the near fatal error of getting a job and moving to California before completing the writing, she admonished me not to become one of those whose "lives were ruined by purebred dogs." I had at one point promised to try to learn from her mistakes and she ruthlessly reminded me of that promise some time ago as a prelude to telling me that her biggest regret was not actually finishing her own theses and that I should not repeat that particular error.

Statistics lessons and assistance were cheerfully provided by Cameron McPhee and Katrina Van Valen Moore. Both of them answered countless questions, sometimes more than once. Katrina, in particular, provided assistance with infinite patience and humor often at the last moment, in spite of the fact that she too is trying to finish her own Ph.D. I am grateful to Ellen Roth for her thoughtful comments on an early draft of several chapters.

My parents have been wondering for quite some time when I was actually going to finish, though thankfully never if I was going to finish. Their unflagging belief in the importance of education has been a touchstone and confidence builder throughout my life.

My friends both locally and afar have been incredibly patient with me during this process, especially these past several months when I have failed to return phone calls or respond to email for weeks at a time, if not longer. It will be good to reconnect with them. It is even better to know that I have friendships that can survive such neglect.

Elizabeth and Dan Hostetter never suspected that they were going to end up living with a borzoi for six months when they offered to take my dogs during the final crunch so that I would stop agonizing about them not getting enough exercise. In the end, they ended up only with Lilly but I am infinitely grateful to them for doing so. In addition to dog sitting ad infinitum, Beth provided computer support, hardware and software upgrades, an airport card and unflappable assistance with inevitable computer crashes at all sorts of odd hours.

Michael Spivey was my friend before he was my advisor and I expect and hope that he will continue to be my friend for many more years. These experiments took shape in conversations held mostly while shooting pool. Although my work is almost completely unrelated to his, Spivey has always listened carefully and offered thoughtful comments. I will never forget the lab meeting in which I first presented the ideas for this research when Spivey silenced half of those asking questions for the remainder of the meeting by informing them that I was studying real children in real classrooms and they needed to think applied.

Finally, but certainly not least, I am grateful to Luke Sollitt, who has been steadfast in his support and encouragement, and judicious in his prodding. Watching him finish his own Ph.D. last summer was no small inspiration for me though I fear that I have required much more patience from him than he did of me. My life is far richer for his being a part of it.

This research was supported, in part, by a National Science Foundation Predoctoral Fellowship. Opinions expressed are those of the author and not necessarily the Foundation.

TABLE OF CONTENTS

CHAPTER ONE: INTRODUCTION (BACKGROUND) 1

Why study word problems? 4

What is mathematical problem solving? 6

Studies with Children 15

Simplifying word problems 20

Textbook content analysis 23

Social and cultural factors 30

Supporting Adult Findings 34

Recall and Recognition of word problems 35

Learning to Solve Word Problems 37

CHAPTER TWO: SOLVING ARITHMETIC WORD PROBLEMS 42

Experiments 42

Experiment 1: Replication/Verification of Problem Type Difficulty 44

Participants 44

Materials 44

Procedure 46

Scoring 48

Results 48

Results by problem 52

Experiment 2: Typical vs Clarified Wording (Change problems) 54

Participants 55

Materials 55

Procedure 56

Results 57

Results by problem 60

Study 3: Consistent vs Inconsistent wording 64

Participants 64

Materials 64

Procedure 65

Results 65

Study 4: Probable and Improbable subtraction 66

Participants 68

Materials 68

Procedure 69

Results 69

CHAPTER THREE: EFFECTS OF ARITHMETIC PRACTICE ON WORD PROBLEM SOLUTION 72

Support for experiments 72

What are canonical and non-canonical problems? 73

Experiment 5: Differential arithmetic practice 75

Participants 75

Materials 76

Canonical Practice 76

Non-Canonical Practice 77
Changes to Posttest 78
Procedure 79
Scoring 81
Overall Results 82
Results of Worksheet Practice 89
Effect of worksheet practice 90

CHAPTER FOUR: CONCLUSIONS/GENERAL DISCUSSION 93

On clarifying word problems 93

On the effects of differential arithmetic practice 95

Other data analyses 98
Identify patterns of errors 98
Analyze children's explanations of what they did 99

Future Directions 100
Structured practice 101
Practicing word problems 103

APPENDIX A: TYPICAL CHANGE PROBLEMS, ORDER A 105

APPENDIX B: CLARIFIED CHANGE PROBLEMS, ORDER A 107

APPENDIX C: COMPARE PROBLEMS, ORDER A 109

APPENDIX D: TYPICAL CHANGE PROBLEMS, POST-TEST, ORDER A 110

APPENDIX E: SAMPLE CANONICAL ARITHMETIC WORKSHEET 112

APPENDIX F: SAMPLE NONCANONICAL ARITHMETIC WORKSHEET 113

REFERENCES 114

LIST OF TABLES

Table 1.1 Classification of whole number addition and subtraction word problems 10

Table 2.1 Percent correct overall by arithmetic operation and gender. 49

Table 2.2 Addition and subtraction problems, percent correct by grade 49

Table 2.3 Mean percent correct by grade and gender. 50

Table 2.4 Mean percent correct for Change and Compare problems by grade 51

Table 2.5 Percentage of students answering correctly on Change problems 53

Table 2.6 Percentage and number of problems correct by computation type and gender 57

Table 2.7 Percentage and number of addition and subtraction problems answered correctly by boys and girls 58

Table 2.8 Mean percent correct for Typical and Clarified problems by grade 59

Table 2.9 Percentage of students answering correctly on Clarified change problems 61

Table 2.10 Percentage of students answering correctly on Typical and Clarified problems 62

Table 2.11 Mean percentage correct for consistent and inconsistent problems by grade 66

Table 3.1 Percent correct and improvement by gender and overall 82

Table 3.2 Percentage improvement by gender and grade 83

Table 3.3 Percentage of students answering correctly on pretest and posttest by problem type 86

Table 3.4 Percent improvement by problem type 87

Table 3.5 Percent improvement by worksheet condition 90

LIST OF FIGURES

Figure 1.1 Consistent and inconsistent language versions of the butter problem (from Mayer and Hegarty, 1996). 35

Figure 1.2 Semantic and literal errors in remembering the butter problem, (Mayer and Hegarty, 1996). 37

Figure 2.1 Examples of Typical and Clarified Change 2 problems 56

Figure 2.2 Examples of consistent and inconsistent problems 65

Figure 2.3 Examples of probable and improbable subtraction problems 69

Figure 2.4 Distribution of correct answers on probable and improbable Compare 6 problems 71

Figure 3.1 Examples of Canonical and Non-Canonical practice problems. 77

Figure 3.2 Examples of pretest and posttest Change 2 problems 79

Figure 3.3 Percent improvement on Typical, Clarified and Compare problems by grade 88

CHAPTER ONE:
INTRODUCTION (BACKGROUND)

It is commonly held belief that the mathematical abilities of children in the United States lags behind that of their age-matched peers in many other industrialized nations, especially Asian countries such as Japan. This impression is supported by the results of recent international assessments of mathematical achievement which tested 13- and 17-year-olds in several countries (Robitaille & Travers, 1992). Why? What is so difficult about mathematics? What sorts of things do children make errors on? Are there systematic patterns to those errors?

Informal conversations with elementary school teachers about which mathematical concepts children find difficult seem to indicate just about everything: fractions; decimals; long division; multi-digit multiplication; word problems; and multi-digit arithmetic, especially when regrouping (i.e., borrowing and/or carrying) is involved (Sherwood, 1997; Blanco, 1997). Single-digit addition and subtraction facts take months to memorize and multi-digit addition and subtraction cause difficulty if regrouping is involved. This difficulty continues to manifest itself when multiplication and division, especially long division, are introduced. When arithmetic moves beyond integers to decimals, fractions and percentages, many children seem to become hopelessly confused. Children also have difficulty with word problems from the time they are introduced into the curriculum through every level of mathematics instruction.

What sorts of errors do children make? The results of the Fourth National Assessment of Educational Progress (NAEP) suggest that children eventually become relatively good at low-level skills such as computation, but remain relatively poor at high-level skills such as problem solving (Mayer & Hegarty, 1996; Kouba, Brown, Carpenter, Lindquist, Silver & Swafford, 1988). Kouba et. al., (1988) attribute the difficulty that young children (third grade students) have with arithmetic to a lack of place value skills. It is easy to see that an incomplete or fragile understanding of how the various digits of a multi-digit numeral relate to one another could lead to errors in regrouping and to understanding decimals later on. Eventually, American children seem to overcome this difficulty; the results of the 4th NAEP indicate that high school students (seventh and eleventh graders) display increasing computational competence. Difficulties with word problems, however, seem to plague children up to and through their college years (Mayer & Hegarty, 1996), and one would presume, into adulthood. Since everyday problems generally are not set out in symbolic form, one must determine the formula necessary prior to solution; this has very real consequences.

This dissertation focuses upon the difficulties young children have understanding word problems, specifically addition and subtraction word problems, for which there is a broad literature. Although children certainly have difficulty with more advanced mathematical concepts such as fractions, algebra, and geometry, there is sufficient evidence that their mastery of basic mathematical concepts is fragile enough that it seems logical to focus on why those basic abilities are difficult and what might be done to improve competence on those tasks. If one tries to build upon a faulty foundation, one can expect to have problems with the building later. Knowledge is no different.

This dissertation discusses reasons that addition and subtraction word problems may be difficult for children to master. The advantage for Asian students has generally been attributed to social and cultural factors, family and culture more supportive and more demanding of academic excellence. Other cross-national studies document that this differential in mathematical ability between American and Asian children exists as early as first-grade or kindergarten (Stevenson, Lee & Stigler, 1986) too early for formalized schooling to be the cause, leading to the suggestion that Japanese students might have an innate cognitive superiority (Lynn, 1982, as cited by Miura, 1987) but Miura and colleagues suggest that the regular structure of number words in many Asian languages may result in a difference in the structure of numerical understanding, in particular, understanding of place value. The ramifications of place value understanding may affect the acquisition of more advanced mathematical concepts later.

Difficulty with word problems probably has other bases. The structure of addition and subtraction word problems has been studied extensively. Models of children's solutions of addition and subtraction word problems have attempted to account for children's difficulties in terms of the surface characteristics or underlying semantic structure of the problem. Experiments involving college students suggest that as the wording of problems becomes more complex (i.e., as the arithmetic solution needed to solve the problem becomes less obvious from a direct reading), solution accuracy decreases. This might be the result of text comprehension difficulties and/or a lack of practice.

Why study word problems?

What makes word problems interesting to study? Word problems are interesting because so many people think word problems are difficult. If one asks a group of people "what's the hardest thing about math?" two answers emerge with frequency: fractions and word problems. Furthermore, people who cite fractions as being the most difficult thing about math will frequently change their minds and agree that word problems are the most difficult thing if they hear them suggested.

It is not just that people think that word problems are hard. There is a considerable body of evidence to support the conclusion that word problems actually are difficult. The results of recent national and international assessments of mathematics achievement (e.g., Dossey, Mullis, Lindquist & Chambers, 1988; LaPointe, Meade & Phillips, 1989; Robitaille & Garden, 1989; Stevenson & Stigler, 1992; Stigler, Lee & Stevenson, 1990) make it clear that although many students perform well on tests of low-level skills such as arithmetic computation, in general, students in the United States tend to perform poorly on tests of high-level skills such as mathematical problem solving. For example, nearly all of the 17-year-olds tested in the 4th NAEP were able to solve basic arithmetic computation problems such as 604-207=?, but nearly all failed to solve multi-step word problems such as (Dossey et. al., 1988):

> Christine borrowed $850 for one year.
> If she paid 12% simple interest on the loan,
> what was the total amount she repaid?

Although many students are able to carry out basic mathematical procedures when problems are presented in symbolic form, they appear to have difficulty applying these procedures when problems are presented in words. In short, these assessments suggest that the difficulty appears to be in understanding word problems rather than in executing arithmetic procedures.

So, what makes word problems so difficult? It's difficult to extract the math (the arithmetic) from the words. Problem wording is often terse. The concise language in which math problems tend to be worded can be somewhat cryptic until one learns the language of them. The correct math problem necessary for solution is not always obvious. Extensive practice at simpler problem types may reinforce solution practices that are not as flexible or general as needed for solving more complex problems

Why should we care? Word problems are important because the real world does not often hand us arithmetic, except perhaps in the case of balancing a checkbook. For the most part, real world problems are story problems. In order to solve them, one must first figure out what the problem is asking and translate that into a mathematical sentence or formula which then needs to be solved. Only after the problem has been identified can we then go about actually solving the resulting arithmetic problem.

What can we do about it? In order to solve the problem we have to understand why children fail to solve problems correctly. Since the difficulty with word problem solution begins early, we will focus on children. Apart from difficulties with computation, which young elementary school children still have, understandably, what are the difficulties they have with the problems? Is it a failure to apply the necessary arithmetic? Do students misunderstand what the problem is asking and if so, how are they interpreting the problem? What do they think the problems say?

What is mathematical problem solving?

A problem exists when a problem solver has a goal but does not know how to reach that goal (Duncker, 1945, as cited in Mayer & Hegarty, 1996). There are three elements involved in the description of a problem – the given state, the goal state, and the allowable operations. Problem solving, or thinking, occurs as the problem solver figures out how to get from the given state to the goal state (i.e., figures out how to solve the problem). Problem solving refers to the processes enabling a problem solver from a state of not knowing how to solve a problem to a state of knowing how to solve it. A problem may be categorized as a mathematical problem whenever a mathematical procedure (i.e., an arithmetic or algebraic procedure) is needed to solve the problem. Thus mathematical problem solving is the cognitive process of figuring out how to solve a mathematical problem that one does not already know how to solve.

According to Riley, Greeno and Heller (1983), a word problem identifies some quantities and describes a relationship among them. Although as adults, we tend to think of word problems as text based problems describing a situation requiring solution, problems vary a great deal in elementary school textbooks from problems displayed entirely with pictures to problems described entirely in words and many intermediary forms combining words and pictures to varying degrees (Stigler, Fuson, Ham & Kim, 1986). Since we are discussing the difficulties that young children have with word problems, we should define word problems in an appropriately broad manner. For the purposes of their analysis, Stigler et. al., (1986) defined a word problem as consisting of 2 or more premises and a question, each presented in verbal form or in an iconic form isomeric to a verbal form.

Word problems seem to be difficult throughout life. Although students for the most part seem to eventually master simpler ones, multi-step problems and problems with difficult language or inconsistent wording are problematic even for college students (Mayer & Hegarty, 1996). On the other hand, students are reasonably good at solving the corresponding equation. This is of concern because life generally doesn't hand us formulas to be solved; we have to determine what the formula is that needs to be solved and then we need to solve it.

In the problem solving literature it is customary to distinguish between two major kinds of problem-solving processes - representation and solution. Representation occurs when a problem solver seeks to understand the problem and solution occurs when a problem solver actually carries out actions needed to solve the problem. There is growing evidence that most problem solvers have more difficulty constructing a useful problem representation than executing a problem solution (Cardelle-Elawar, 1992; Cummins, Kintsch, Reisser & Weimer, 1988; DeCorte, Verschaffel & DeWin, 1985). Mayer (1985, 1992, 1994) has proposed four main component processes in mathematical problem solving: translating, integrating, planning, and executing. Translating involves constructing a mental representation of each statement in the problem. Integrating involves constructing a mental representation of the situation described in the problem. Planning involves devising a plan for how to solve the problem. Executing involves carrying out the plan, including computations. The first two processes, translation and integration, are involved in problem representation. Planning is a natural product of problem representation. Students frequently correctly devise and carry out computational plans based on an incorrect representation of the problem. Mayer and Hegarty (1996) maintain that an important key to mathematical problem solving rests in the processes by which students seek to understand math problems. They contend that the major creative work in solving word problems rests in understanding what the problem means; carrying out a solution plan follows naturally from the problem solver's representation of the problem.

This dissertation concentrates on addition and subtraction word problems for several reasons. First, there is a generally accepted categorization schema that has been worked out. Next, it is generally accepted that competence in addition and subtraction skills is necessary before attempting to teach multiplication and division. Multiplication and division word problems have also been studied, but not to the extent of addition and subtraction word problems. It also seems logical to study the most basic mathematical concepts children have difficulty with as understanding these difficulties may shed light on what is difficult about concepts introduced later, or more hopefully, a better grounding in those basics may improve performance on other skills.

Most current research attempts to examine the processes that children use to solve arithmetic problems (Carpenter & Moser, 1982). The development of basic addition and subtraction concepts is described in terms of levels of increasingly sophisticated and efficient problem solving strategies. Attempts to characterize word problems have focused either on syntactic variables, the semantic structure of the problem, or some combination of the two. Syntactic variables such as the number of words in a problem, the sequence of information, and the presence of words that cue a particular operation do significantly affect problem difficulty (see Carpenter & Moser, 1982 for a review of the research on problem difficulty), but most of the evidence available suggests that the semantic structure of a problem is much more important than syntax in determining the processes that children use in their solutions (Carpenter, 1985; Carpenter, Ansell, Franke, Fennema & Weisbeck, 1993). The semantic structure of addition and subtraction word problems has been classified and described in a number of different ways. Many different terms have been given to identified situations, but there is considerable overlap in the situations used in most category systems. Riley et al., (1983) introduced a classification scheme for simple addition and subtraction word problems that distinguishes four broad categories of problems based on the semantic structure of the problems: Change, Combine, Compare and Equalize problems (see Table 1.1).

Change problems refer to dynamic situations in which some event changes the value of a quantity. Combine problems refer to situations involving two quantities that are considered either separately or in combination. Compare problems involve two quantities that are compared and the difference between them. Equalize problems are a hybrid of Compare and Change problems (Carpenter & Moser, 1992).

Fuson (1992a, b) notes that although most category systems collapse the static/dynamic distinction into the binary/unary distinction, yielding only static binary Combine and Compare problems and dynamic Change problems, dynamic binary forms of Combine and Compare problems can be constructed and are frequently easier to solve than static forms because the actions in the problems cue solution procedures. Equalize problems are active binary Compare problems in which the difference between two quantities is expressed as unary change actions rather than as a static state as in Compare problems. Dynamic Combine problems can be created by making the combining explicit rather than implicit using class inclusion terms or with words such as “altogether.”

Table 1.1 Classification of whole number addition and subtraction word problems

CHANGE	
Join or Add To	**Separate or Take From**
Result Unknown Pete had 3 apples. Ann gave him 5 more apples. How many apples does Pete have now?	Result Unknown Joe had 8 marbles. Then he gave 5 marbles to Tom. How many marbles does Joe have now?
Change Unknown Kathy had 5 pencils. How many more pencils does she need so she has 7 pencils altogether?	Change Unknown Fred had 11 pieces of candy. He lost some of the pieces. Now he has 4 pieces of candy. How many pieces of candy did Fred lose?
Start Unknown Bob got 2 cookies. Now he has 5 cookies. How many cookies did Bob have in the beginning?	Start Unknown Karen had some word problems. She used 22 of them in this table. She still has 79 word problems. How many word problems did she have to start with?

COMBINE physically	EQUALIZE	
	Join	**Separate**
Combine value Unknown (Join)	Difference Unknown	Difference Unknown
Sara has 6 sugar donuts and 9 plain donuts. Then she puts them all on a plate. How many donuts are there on the plate?	Susan has 8 marbles. Fred has 5 marbles. How many more marbles does Fred have to get to have as many marbles as Susan has?	Jane has 7 dolls. Ann has 3 dolls. How many dolls does Jane have to lose to have as many as Ann?
Subset Unknown (Separate)	Compared Quantity Unknown	Compared Quantity Unknown
Joe and Tom have 8 marbles when they put all their marbles together. Joe has 3 marbles. How many marbles does Tom have?	There were 6 boys on the soccer team. Two more boys joined the team. Now there is the same number of boys as girls on the team. How many girls are on the team?	There were 11 glasses on the table. I put 4 of them away so there would be the same number of glasses as plates on the table. How many plates were on the table?

Table 1.1 (Continued)

	<u>Referent Unknown</u> Connie has 13 marbles. If Jim wins 5 marbles, he will have the same number of marbles as Connie. How many marbles does Jim have?	<u>Referent Unknown</u> There were some girls in the dancing group. Four of them sat down so each boy would have a partner. There are 7 boys in the dancing group. How many girls are in the dancing group?
<u>COMBINE conceptually</u>		**<u>COMPARE</u>**
	Join	**Separate**
<u>Combine value Unknown (Join)</u> There are 6 boys and 8 girls on the soccer team. How many children are on the team?	<u>Difference Unknown</u> Joe has 3 balloons. His sister Connie has 5 balloons. How many more balloons does Connie have than Joe?	<u>Difference Unknown</u> Janice has 8 sticks of gum. Tom has 2 sticks of gum. Tom has how many sticks less than Janice?
<u>Subset Unknown (Separate)</u> Brian has 14 flowers. Eight of them are red and the rest are yellow. How many yellow flowers does Brian have?	<u>Compared Quantity Unknown</u> Luis has 6 pet fish. Carla has 2 more fish than Luis. How many fish does Carla have?	<u>Compared Quantity Unknown</u> The milkman brought on Sunday 11 bottles of milk and on Monday he brought 4 bottles less. How many bottles did he bring on Sunday?
	<u>Referent Unknown</u> Maxine has 9 sweaters. She has 5 sweaters more than Sue. How many sweaters does Sue have?	<u>Referent Unknown</u> Jim has 5 marbles. He has 8 fewer marbles than Connie. How many marbles does Connie have?

Note: This table is adapted from Fuson (1992a). The problems are taken from a variety of sources and are presented in order of difficulty with problems becoming more difficult from left to right and from top to bottom. The easiest problems are thus at the top left of the table and most difficult problems at the bottom right.

Each of these categories can be further subdivided into distinct problem types depending on the identity of the unknown. In each category there are three types of information. In change problems, the unknown may be the start, result or change set. Similarly for Compare problems, the unknown quantity may be the difference, the compared quantity or the referent. The unknown quantity in Equalize problems can be varied to produce three distinct types although Equalize problems are not commonly found in the research literature or in most mathematics programs (Carpenter & Moser, 1982, Stigler, Fuson, Ham & Kim 1986). In combine problems the unknown is either the combined set or one of the subsets.

For Change, Equalize, and Compare problems, further distinctions can be made depending on the direction of the event (i.e., increase or decrease) or the relationship (i.e., more or less). In their research, Carpenter and Moser often refer to additive and subtractive Change problems as Join and Separate problems, respectively.

Robust research evidence is now available which shows the psychological significance of the semantic classification of word problems. Word problems that can be solved by the same arithmetic operation but differ with respect to their underlying semantic structure have very different degrees of difficulty (DeCorte & Verschaffel, 1991).

A number of models have been developed to simulate young children's understanding and solution of simple word problems concerned with the exchange, combination, and comparison of sets (Briars & Larkin, 1984; Cummins et al., 1986; Kintsch & Greeno, 1985; Reusser, 1989, 1990; Riley & Greeno, 1988; Riley et al., 1983). Stern and Lehrndorfer (1992) point out that a common feature of all of these models is that the fit between the model predictions and the empirical data is better for Change and Combine problems than for Compare problems. None of the models can explain why Compare problems are so difficult, nor why different kinds of Compare problems differ in difficulty. Evidence points to non-mathematical factors such as language understanding, text comprehension and situational understanding factors.

Mayer and Hegarty (1996) suggest that there are two strategies for dealing with mathematical story problems. When confronted with a mathematical story problem, some people use a direct translation strategy – they seem to begin by selecting numbers from the problem and preparing to perform arithmetic operations on them. Other people use what Mayer and Hegarty call a problem model strategy – they try to understand the situation being described in the problem and devise a solution plan based on their representation of the situation. Mayer and Hegarty characterize the direct translation strategy as a short-cut heuristic approach that emphasizes computation, in contrast with the problem model approach as an in-depth rational approach based on problem understanding. The direct translation strategy emphasizes quantitative reasoning - computing a numerical answer - whereas the problem model strategy emphasizes qualitative reasoning, understanding the relations among the variables in the problem. Stigler, Lee and Stevenson (1990, p. 15) summarize this short-cut approach as "compute first and think later" because the problem solver engages in quantitative reasoning prior to qualitative reasoning (Mayer, Lewis & Hegarty 1992).

The direct translation strategy is familiar as the method of choice for less successful problem solvers in several research literatures. Cross-national research on mathematical problem solving reveals that American children are more likely than Japanese children to engage in short cut approaches to story problems and that instruction in US schools is more likely than instruction in Japanese schools to emphasize computing correct numerical answers at the expense of understanding the problem (Stevenson & Stigler, 1992; Stigler et al., 1990). Similarly, research on expertise reveals that novices are more likely to focus on computing a quantitative answer to a story problem than experts who are more likely initially to rely on a qualitative understanding of the problem before seeking a solution in quantitative terms (Chi, Glaser & Farr, 1988, Smith 1991, Sternberg & Frensch, 1991). The direct translation strategy does make minimal demands on memory and does not depend on extensive knowledge of problem types, but it frequently leads to incorrect answers (Hegarty, Mayer & Green 1992; Lewis & Mayer, 1987; Mayer, Lewis & Hegarty 1992; Verschaffel, DeCorte & Pauwels 1992).

In contrast to the direct translation strategy, the problem model strategy consists of constructing a qualitative understanding of the problem situation before attempting to carry out arithmetic computations. The problem solver begins by constructing an internal representation of each of the individual statements in the problem and seeks to understand the general situation described in the problem before constructing a plan for solving the problem. These three components – local understanding of the problem statements, global understanding of the problem situation, and construction of a solution plan – constitute three major components of mathematical problem solving according to Mayer (1985, 1992).

Understanding of a problem has long been recognized as one of the premier skills required for successful mathematical problem solving (Cummins, Kintsch, Reisser & Weimer, 1988; Greeno, 1987; Mayer, 1985, 1991; Mayer, Larkin & Kadane, 1984; Polya, 1965; Wertheimer, 1959). Problem understanding occurs when a problem solver converts the words of the problem into an internal mental representation of the problem situation.

Studies with Children

Given the difficulty that college students have with word problems, it seems reasonable to suggest that perhaps young children are simply missing the cognitive competence necessary to deal with them. Constructing a model or representation of a problem situation is one of the most fundamental processes of problem solving. Many problems can be solved by representing directly the critical features of the problem situation. Modeling, it turns out, is also a relatively natural process for young children. An extensive body of research documents that even prior to receiving formal instruction in arithmetic (maybe especially before) young children are able to solve a variety of different types of addition and subtraction word problems by directly modeling the different actions and relationships described in the problems with counters (Carpenter, 1985; Fuson, 1992a, b).

On the other hand, some of the most compelling exhibitions of problem solving deficiencies in older children appear to have occurred because the students did not attend to what appear to be obvious features of the problem situations. For example, in one frequently cited item from the third national mathematics assessment of the NAEP (1983), students were asked to find the number of buses required to transport 1128 soldiers if 36 soldiers could ride in each bus. Although nearly three quarters of the 13-year-olds tested recognized that division was required to solve the problem, only about one third of them rounded the quotient up to the next largest whole number to account for the fact that the answer must be a whole number of buses. Most students either reported a fractional number of buses or rounded down, leaving 12 soldiers stranded without transportation. This is one of many examples suggesting that many students abandon a fundamentally sound and powerful general problem-solving approach for the mechanical application of arithmetic and algebraic skills. It appears that if older children would simply apply some of the intuitive analytical modeling skills exhibited by young children to analyze problem situations, they would avoid some of their more glaring problem solving errors. A fundamental issue would seem to be how to help children build upon and extend the intuitive modeling skills that they apply to basic problems as young children.

Carpenter et. al. (1993) focused not on instruction but on the problem solving processes of children. They did not address exactly how instruction should be designed to accomplish this task as they claimed to be particularly concerned with how an analytical framework based on the notion of problem solving as modeling explained children's strategies for solving problems. Although there is some variability in children's performance depending on the nature of the action or relationships in different problems, by the first grade, most children can solve a variety of problems by directly modeling the action or relationships described in them. There are two accounts of the cognitive mechanisms involved in these situations that differ in fundamental ways. Riley et al., (1983; Riley & Greeno, 1988) propose that children's ability to solve simple addition and subtraction problems depends on the availability of specific problem schemata for understanding the various semantic relationships in the problems. Briars and Larkin (1984), on the other hand, propose an analysis that, at the most basic level, does not include separate schemata for representing different classes of problems. Problems are mapped directly onto the action schemata required to solve the problem. In other words, Riley and her associates hypothesize that specific knowledge about additive structures is required to solve basic addition and subtraction problems, whereas Briars and Larkin propose that children's initial solutions can be accounted for essentially in terms of the actions required to model the action in the problem. Other accounts of the processes involved in solving word problems such as the linguistic analysis of children's difficulty in translating natural language statements into action on and relationships among sets (e.g., Cummins, 1991; Cummins, Kintsch, Reisser & Weimer, 1988) generally build on the basic semantic analyses and must ultimately deal with the issue of whether or not it is necessary to hypothesize specific knowledge of additive structures to account for children's behavior.

Studies have shown that giving children experience with addition and subtraction problem types that are not typically a part of the primary mathematics curriculum can significantly improve performance and reduce the discrepancy between problems that are considered relatively easy and certain problems that are generally considered more difficult.

Seventy Kindergarten children who had spent the year solving a variety of basic word problems were individually interviewed as they solved addition, subtraction, multiplication, division, multi-step and non-routine word problems. Counters and paper and pencil were available and children were told that they could use any of those materials to help them solve the problems. Problems were reread as many times as the child wished. The kindergarten children tested showed a remarkable degree of success in solving word problems. Nearly half of the children used a valid strategy for all of the problems administered, and almost two-thirds correctly solved at least seven of the nine problems. Almost all of the children used a valid strategy for the most basic subtraction and multiplication problems and over half of the children were successful even on the most difficult (non-routine) problem. It is interesting to compare the result of the division with remainder problem in which the kindergarten children were asked to determine the number of cars needed to take 19 children to the circus if 5 children could ride in each car with the related National Assessment item. Although the numbers used in the two problems were vastly different, most of the kindergarten children had no difficulty deciding how to deal with the remainder, unlike the 13-year-olds. In fact, almost as many children correctly solved the division with remainder problem as solved the division problem without remainder.

Children can solve a wide range of word problems, including problems involving multiplication and division, much earlier than has generally been presumed. American textbooks typically include a narrow range of addition and subtraction problems in the primary grades (Stigler, Fuson, Ham & Kim, 1986), and multiplication and division problems are not introduced until late in the second grade. The results of this study suggest that much more challenging problems involving a range of operations can be introduced in the early primary grades.

With only a few exceptions, the children's strategies for solving the problems could be characterized as directly representing or modeling the action or relationships described in the problems. Although instruction did encourage the use of modeling to solve problems, the children in this study successfully model problems that differed from the problems they had seen in class, suggesting they can apply this ability to a reasonably broad range of problems.

Although these findings are more consistent with the more general analysis of problem solving proposed by Briars and Larkin (1984), they do not conclusively demonstrate that specific multiplication and division schemata are not required for successful solving of multiplication and division problems as hypothesized by Riley, Greeno and Heller (1983). The results do suggest however that if such specific schemata are required, they are already sufficiently well developed in many kindergarten children that they can solve multiplication and division problems by representing the action and relationships in the problems. Perhaps at a more fine grained level of analysis, specific schemata are necessary in order to account for children's performance, but describing performance in terms of modeling provides a parsimonious and coherent way of thinking about children's mathematical problem solving that is relatively straightforward and accessible to students and teachers alike (Carpenter, Fennema & Franke, 1992 as cited in Carpenter, Ansell, Franke, Fennema & Weisbeck, 1993). This conception of problem solving as modeling could provide a unifying framework for thinking about problem solving in the primary grades. It seems to be a basic process that comes relatively naturally to most primary grade children, If we could help children build upon and extend the intuitive modeling skills that they apply to basic problems as young children we would have accomplished a great deal by way of developing problem solving abilities in children in the primary grade. Modeling provides a framework in which problem solving becomes a sense making activity and may have an impact on children's conceptions of problem solving and of themselves as problem solvers (need citation for Frank's example of groups doing better or worse depending on whether they are told that the group of which they are a member generally does well or poorly in such tasks).

Verschaffel (1984, as reported in DeCorte & Verschaffel, 1991) observed a tendency among first-grade children given a series of addition and subtraction word problems to solve the subtraction problems by applying the solution strategy corresponding most closely to the semantic structure of the problem.

So what happens to this powerful general problem solving ability? It goes away. Why? That hasn't been clearly answered. I believe it is because we present children with simple problems that don't require such a general solution procedure and they learn to extract a simpler process of extracting a number sentence (usually canonical) from the text. Since canonical sentences are over learned from their sheer volume, children do not get practice at solving more complex problem forms until it either does not occur to them to try, or they become much more likely to make errors because such problems are infrequent and unpracticed.

The relationship between the semantic structure of simple addition and subtraction word problems on the one hand, and children's solution strategies on the other, holds not only for children solving problems with the help of concrete objects such as fingers or blocks, but also for those applying counting strategies, whether verbal, based on counting forward or backward, or mental, based on recalled number facts (DeCorte & Verschaffel, 1987).

Simplifying word problems

Although semantic structure does appear to be a major factor determining problem solution, recent research has made it clear that other task characteristics can also significantly alter children's performance and strategies on verbal problems. Two that have been investigated include the degree to which the underlying semantic structure is made explicit in the problem text, and the order of presentation of the given numbers.

When conducting individual interviews with first-grade children, Verschaffel (1984, as cited in DeCorte & Verschaffel, 1991) observed that some children who could not solve a standard Combine 2 problem (e.g., Ann and Tom have 8 books altogether; Ann has 5 books, how many books does Tom have?), were able to solve a reworded version of the problem in which the surface structure made the semantic relations more obvious (e.g., Ann and Tom have 8 books altogether; 5 of these books belong to Ann and the rest belong to Tom; How many books does Tom have?). Carpenter (1985) also showed that subtle aspects of the formulation of the problem, such as the tenses of the verbs in the problem text, may be responsible for observed differences in difficulty between variants of Change problems.

Hudson (1983) demonstrated that kindergarten children are much better at simplified Compare problems than standard ones. Young children presented with a picture of 5 birds and 4 worms performed much worse when asked the more standard question “How many more bird than worms are there?” than the alternative “Suppose the birds all race over and try to get a worm; how many birds won’t get a worm?”. In the latter case, most of the children appeared to use a matching strategy to solve the problem. Compare problems have been broadly found to be the most difficult type of word problem, but Hudson’s data suggest that childrens’ difficulties on Compare problems are influenced by the formulation of the problem.

Based on these findings, DeCorte, Verschaffel and DeWin (1985) systematically tested the hypothesis that rewording simple addition and subtraction word problems in such a way that the semantic relations are made more explicit without affecting the semantic structure of the problem would facilitate the solution of these problems by young elementary school children.

Two sets of six rather difficult word problems were formulated - two each of Combine 2, Change 5, and Compare 1 problems - were administered near the end of the school year to a group of 89 first- and 84 second- graders. In one set, the problems were stated in the usual 'condensed' form; in the other set, they were reformulated to make the semantic relations more explicit. The reworded problems were solved significantly better than the standard problems.

DeCorte and Verschaffel (1991) hypothesize that the process of constructing a representation of the problem is a complex interaction of top down and bottom up processes - that is, the processing of verbal (textual) input as well as the problem solvers semantic schemes both contribute to the construction of a representation. For less able and inexperienced children, semantic schemes are not very well developed, so they depend more on text driven processing to construct an appropriate problem representation.

The sequence of the numbers (and information) in the problem text also affects children's solution processes. Verschaffel (1984) found that children solved Combine 2 problems either by adding on (when using concrete objects) or by counting up (when using verbal counting strategies) from the smaller given number. On the other hand, Carpenter and Moser (1984) reported that children in their study tended to either separate from or counting down from the larger given number. A closer examination of the problems used in both studies reveals that in the Verschaffel problem, the larger number was mentioned first (e.g., Pete has 3 apples; Ann also has some apples; Pete and Ann have 9 apples altogether; how many apples does Ann have?), and in the Carpenter and Moser problem, the larger number was given first (e.g., There are 6 children on the playground; 4 are boys and the rest are girls; how many girls are on the playground?), suggesting that the strategies young children use to solve addition and subtraction word problems depend not only on the semantic structure of the problem, but also on the sequence of the given numbers in the task.

While rewording problems to make the semantic structure more explicit may assist younger children in solving a greater range of problems, we must not forget that in the long run, children must learn to solve more tersely worded problems and problems with more complicated wording.

One wonders if problems can also be made more difficult by altering their wording. For example, one would expect that changing the order of information in a Change 1 problem to "Joe gave Stephanie 4 books. Before that Stephanie had 7 books. How many books does Stephanie have now?" should increase the difficulty of the problem for young children.

Textbook content analysis

Fuson, Stigler & Bartsch (1988) studied the grade placement of addition and subtraction topics in elementary school textbooks in mainland China, Japan, the Soviet Union, Taiwan and the United States. Mainland China, the Soviet Union and Taiwan all had a national curriculum which used a single textbook series for the entire country. In Japan, the math curriculum is set by the Ministry of Education (Mombusho) and although there are several textbook series, they all adhere to the placement of topics specified by the ministry (Stevenson, Lummis, Lee & Stigler, 1990). There is a high degree of uniformity in the grade placement of these topics in China, Japan, the Soviet Union and Taiwan, and substantial difference between the placements for those countries and for the United States. Single-digit addition and subtraction problems (i.e., the addition of single digits or the subtraction of single digits yielding single digit answers) appear and disappear earlier in the textbooks of other countries than in those of the US. Both the simplest and the most difficult multi-digit addition and subtraction appear from one to three years earlier in those textbooks than in US textbooks.

Stevenson et at. (1990) and Fuson et al. (1986) found that the only topics that appeared earlier in US textbooks than in Japanese textbooks were ratio and proportion, problem solving, fractions, and weight.

Sugiyama (1987, as cited in Robitaille & Travers, 1992) concluded that the word problems in Japanese textbooks were more difficult than those found in American textbooks. Problems for grades 7 and 8 in the US were found in grade 5 in Japan. Both Stevenson et al. (1990) and Bartsch et al. (1986) found that concepts tended to be introduced up to a year earlier in secondary school textbooks in Japan than they were in US textbooks. Furthermore, there was much more repetition in American books. Over 70% of concepts were repeated at least once after their initial introduction, almost 25% were repeated twice and 10% were repeated 3 times. In Japan, 38% of the topics were reviewed once and only 6% more than once.

Secondary textbooks in the US all tend to be much longer than those used in Japan. American textbooks ranged from 400 to 856 pages with an average of 540 pages while Japanese textbooks were no longer than 230 pages and averaged 178. In addition, more of the problems in Japanese textbooks tended to be complex.

Stigler, Lee, Lucker and Stevenson (1982) provide the only attempt I am aware of to analyze the mathematical performance of children relative to what they have been taught. They concluded that Taiwanese children performed more effectively than their American counterparts.

There are really two questions that need to be answered. The first is whether children in the US lag behind their grade mates in other countries on tests of mathematical skill. The cross-national studies done thus far suggest that this is indeed the case. Another question that has yet to be answered is whether American children are as effective at learning what they are taught as other children. Stigler, Fuson, Ham and Kim (1986) attempted to answer this but their conclusion that Taiwanese children are more effective at learning than American children was based partly on a conclusion that the textbooks used by the two countries were similar. If anything, the Taiwanese textbook presented material more slowly. However, only one US textbook was analyzed and in a subsequent study by Fuson, Stigler and Bartsch (1986) which examined five popular US textbook series, American textbooks were found on average to present material later than the Taiwanese national textbook series.

The research on children's solutions of simple addition and subtraction word problems have made it clear that the ease with which children solve a particular problem varies according to the semantic structure of the problem, the position of the unknown quantity, and the precise way in which the problem is worded. Although it is also logical to expect that the frequency with which children are exposed to problems of different types should relate to the ease with which problems are solved, surprisingly little research has addressed the question of how problems are distributed throughout the elementary mathematics curriculum or the effect that this distribution might have on children's performance on these problems (Stigler et al., 1986).

Stigler et al., (1986) analyzed the word problems in grades one through three of four widely used textbook series in the United States, and compared the number, range, and organization of problems with those in the textbook series mandated by the Soviet government. They found that the Soviet series had more problems (493) across all three grade levels than any of the American texts (which ranged from 328 to 430). The Soviet series also included more two-step problems than any of the American series. Across all four of the American textbook series, only 7% of the problems were two-step problems while 44% of the problems in the Soviet series were two-step problems. Furthermore, many of the few two-step problems in the American textbooks were designated as special "challenge" problems not necessarily targeted to all children. The distribution of problems also varied considerably from series to series. The number of both one- and two-step problems rises precipitously after first grade in three of the American texts, and gradually in the fourth. The number of both types of problems drops precipitously in the Soviet series. In the third grade, the bulk of the word problems in the Soviet series involve multiplication and division. Overall, the Soviet first-grade text contained between three and ten times as many addition and subtraction word problems as the American series.

In general, Stigler et al., (1986) found that Soviet textbooks presented a fairly even distribution of problem types averaged over the three grades, while the American textbooks showed a marked irregularity in the frequency of occurrence of different problems types. Only one type of compare problem (type 1) is presented with any frequency in the American textbooks while the Soviet textbook presents approximately equal numbers of all 6 types. Likewise, the Soviet textbook presents a fairly equal distribution of all 6 kinds of change problems while the American texts present predominately 2 of the 6 types (Change 1 and Change 2). Of the 2 different types of combine problems, American texts present, at best, twice as many missing whole problems as missing part problems (and at worst, ten times as many). As one might guess by now, the Soviet text presents equal numbers of both types of Combine problems.

In the Soviet texts, the most frequent problem type comprised only 9% of the total, while in the American series, the most frequent type comprised nearly a third of the problems. In addition, the three most numerous problem types in the Soviet text were all two-step problems. In the American textbooks, the three most numerous problems types are all one-step problems.

All of the high frequency problems in the American textbooks have semantic structure equations that are identical to their solution procedure equations, what Mayer and his colleagues would call consistently worded problems. The vast bulk of all problems in the American texts are of this simplest form: ones in which the arithmetic solution procedure directly parallels the semantic structure of the problem. These most frequent examples are by far the easiest for American children to solve according to the literature (Carpenter & Moser, 1983, 1984; Riley et al., 1983). Stigler, Fuson, Ham and Kim (1986) conclude that there is a clear bias in the American textbooks toward presenting the problems that American children find easiest to solve, but it is equally possible that American children find these types of problems easiest to solve because they have had the most practice solving them.

In addition to the frequency with which different types of problems were presented, Stigler, Fuson, Ham and Kim (1986) also examined the way in which problems were sequenced. They found that the Soviet textbook series presented far more variability, presenting both a larger variety of problems types in any given group of 10 problems, and greater variability in the ordering of the problems. American textbooks had a tendency to group like problems together. Finally, word problems tended to be distributed quite evenly throughout the Soviet texts while they tended to be grouped together in the American series.

It seems clear that the frequency of exposure might impact the relative difficulty of problems of different types. DeCorte, Verschaffel, Janssens, & Joillet (1984, as cited in DeCorte & Verschaffel, 1991) report that an analysis of the addition and subtraction word problems found in six first-grade Belgian textbooks reveals a similar restrictedness in the range and type of problems presented to that found in the American texts. There was a preponderance of Change 1 and 2 and Combine 1 problems. Only one of the texts presented a variety of Compare problems; in three of the texts there were no Compare problems, and in two there were very few. There is evidence that exposure to uncommon problem types improves performance (see Carpenter, Ansell, Franke, Fennema & Weisbeck, 1993). DeCorte, Verschaffel and DeWin (1985) observed that the word problems in Flemish elementary math textbooks are usually stated very briefly, sometimes even ambiguously, for someone unfamiliar with the standard problem situations such as a young child.

Children spontaneously use a wide variety of informal solution strategies to solve word problems (DeCorte & Verschaffel, 1987, Carpenter & Moser, 1982, 1984). Effective instruction builds on existing knowledge and skills. The errors children make on word problems are remarkably systematic. Like the difficulties adults exhibit, they result from misconceptions of the problem situation. Most researchers argue that such misconceptions are due to an insufficient mastery of the semantic schemes underlying the problem. Both the syntactic structure and the mathematical structure of the problem contribute to understanding difficulties. For example, DeCorte and Verschaffel (1985) found that some children misinterpreted the sentence "Pete and Ann have 9 apples altogether" to mean "Pete and Ann each possess 9 apples" (see DeCorte & Verschaffel, 1985 and Riley et al., 1983 for additional examples). Too often researchers interpret errors as being the result of trial and error behavior or sloppiness, or they ignore errors as 'uninterpretable. " Findings like the above strongly suggest that children make the errors they do because they interpret the problem differently than the adult who wrote it intended. In my own research on preschool children's concepts of numbers, I asked children to count a number of objects and observed on several occasions children making comments to the effect of "What number is this ?" while indicating a particular object, suggesting that at that age, they still seemed to think of numbers as an alternative label for an object, and had not yet fully grasped then flexible nature of numerosity.

Children clearly begin with a variety of flexible strategies for solving a variety of arithmetic problems. Carpenter et al. (1993) have demonstrated that even kindergarten children are able to apply these strategies to multiplication and division problems, and even two-step and irregular word problems. Why they abandon these strategies is not entirely clear, although evidence points toward the fact that such strategies are unnecessary for solving the vast majority of problems that children encounter in the elementary mathematics curriculum. What is clear from the systematic errors made by most children on simple addition and subtraction problems is that they misinterpret these problems. While simplifying problems for very young children in order to make the semantic structure of the problem more obvious, it does not solve the problem. Analyses of textbooks have demonstrated that American textbooks are populated to an overwhelming degree by the simplest types of problems, which are in turn the problems that American children find easiest to solve. Although empirical research has yet to demonstrate the efficacy of presenting a broad range of problems, and such research should be done, there are sufficient hints from international assessments of mathematics achievement and variability in textbooks that teaching experiments in the United States are warranted.

Social and cultural factors

Differences in the amount of class time spent on math, variations in teaching practices and personal characteristics of the students have all been proposed as factors contributing to differences in performance. Stevenson et. al. (1990) reported that Japanese and Taiwanese first-grade students spend more hours per week on math than do US children. Teachers in Japan may spend an entire 40-45 minute class period on just one or two arithmetic problems and they often use student errors as examples for analysis in their teaching (Stigler & Perry, 1988, as cited in Miura, Okamoto, Kim, Steere and Fayol, 1993). Hess and Azuma (1991) have suggested that Japanese children bring personal characteristics to the classroom learning situation that make the particularly receptive to learning.

Stevenson, Lee and Stigler (1986) attempted to address the complaint that comparative studies of children's scholastic achievement have been hindered by the lack of culturally fair, interesting and psychometrically sound tests and research materials. In order to test children in Taiwan, Japan and the United States, a team of bilingual researchers from each culture constructed tests and other research instruments with the aim of eliminating as much of the cultural bias as possible. Mathematics tests were based on the content of the textbooks. The test for kindergarten children contained items assessing basic concepts and operations included in the curricula from kindergarten through third grade, that for elementary school children (first and fifth graders) contained items derived from the concepts and skills appearing in the curricula through grade 6.

American children scored lower on the mathematics achievement tests than Japanese children at all three grades, and lower than Chinese children at grades 1 and 5. Among the 100 top scoring individuals on the math test at the first grade level, there were only 15 American children and only one American child appeared in the top 100 scorers at the fifth grade level. More than half the children scoring in the lowest 100 scores at the first and fifth grade levels were American children (58 in grade 1 and 67 in grade 5). The low level of performance of American children was not due to a few exceptionally low scoring classrooms nor to a particular area of weakness; they were as ineffective in calculating as in solving word problems.

Based on extensive observations, American first-grade children were engaged in academic activities a smaller percentage of the time (69.8) than were Chinese (85.1%) and Japanese (79.2%) children. By the fifth grade, these differences were even greater than at lower grades: American children spent 64.5 percent of their classroom time involved in academic activities where Chinese children spent 91.5% and Japanese children spent 87.4 percent. They estimate this to mean 19.6 hours per week (64.5 percent of the 30.4 hours the American children spent in school, less than half of the estimated 40.4 hours (91.5 percent of the 44.1 hours that Chinese children spent in school), and less than two-thirds of the 32.6 hours (87.4 percent of 37.3 hours Japanese children attend school).

In both grades 1 and 5, American children spent less than 20% of their time studying mathematics, less than the percentage for either Chinese or Japanese children. In the fifth grade, language arts (including reading) and mathematics occupied approximately equal amounts of time in both Chinese and Japanese classrooms. American children spent more than twice as much time (40%) on language arts as they did on mathematics (17%). American teachers spent proportionately much less time imparting information (21%) than did the Chinese (58%) or Japanese (33) teachers. This means American fifth graders were receiving information approximately 6 hours per week (0.21 times 30 hours) compare with estimates of 26 hours for Chinese children and 12 hours for Japanese children. American children were also absent from the classroom more frequently than their counterparts when a child was know to be at school. This almost never occurred (0.2%) in Japanese and Taiwanese classrooms.

These differences become even more profound when extended over the course of the school year. Chinese and Japanese children have fewer holidays and a longer school year (240 days) than do American children (178 days). There are enormous differences in the amount of schooling young children receive in the three countries.

At the time the data were collected, both Chinese and Japanese children spent a half day at school on Saturdays as well. This was reduced at least in the Japanese schools to every other Saturday in the mid-80s, and has recently been discontinued altogether. Since it is unlikely that the Japanese school year has been extended to compensate, it will be interesting to see if this reduction in schooling will affect the mathematics achievement scores of Japanese students. Of course, it is possible that Japanese students will simply spend compensatory time at Juku or cram school and no differences may be noted.

Neither American parents nor teachers of elementary school in the US tend to believe that homework is of much value. American children spend less time on homework than do Japanese children and both groups spend much less time on homework than do Chinese children. American first-grade students spent an average of 14 minutes a day (as estimated by their mothers) on homework while Chinese first-grade students spent an average of 77 minutes per day and Japanese children spent 37. This increased to 46, 114, and 57 minutes per day respectively for American, Chinese and Japanese fifth-grade students. On weekends American children studied even less (7 minutes on Saturday and 11 on Sunday) while Chinese and Japanese children studied a comparable amount to weekdays (83 minutes on Saturdays and 73 minutes on Sundays for Chinese students, 37 and 29 minutes respectively for Japanese students in addition to a half day of school on Saturday). Nearly all of the Japanese (98%) and Chinese (95%) fifth grade students had a desk at home but only 63 percent of American first-grade children had a desk, a statistic Stigler et al. (1982) believe to be indicative of parental concern about schoolwork.

Less than a third of the parents of American fifth graders bought workbooks in mathematics for their children, half as many as the Chinese and Japanese parents. Most American children indicated that they disliked doing homework. Most Chinese children indicated that they enjoyed doing homework; the reaction of the Japanese children were mixed. Most American mothers thought that the amount of homework assigned to their children was “just right”, as did the Chinese and Japanese mothers whose children were assigned far greater amounts of homework (Hess, Chang & McDevitt, 1987). American mothers were overwhelmingly pleased with the job the school was doing teaching their children. Over 90% of the American mothers thought that the school was doing a good or excellent job. Less than half of the Chinese or Japanese mothers rated the school their child was attending so highly.

Supporting Adult Findings

Mayer and Hegarty and associates have been carrying out a program of research that uses a variety of approaches to examine how experienced students read (Hegarty, Mayer & Monk, 1995; Hegarty, Mayer & Green, 1992), remember (Hegarty, et. al., 1995; Mayer, 1982), and learn to solve (Lewis & Mayer, 1987, Lewis, 1989) word problems. Much of their work involves two-step compare problems in which the first step requires addition or subtraction and the second step involves multiplication or division. The relational term is either consistent or inconsistent with the operation required for correct solution. In consistent language problems, the required operation for the first step is primed by the key word (e.g., if the required operation was addition, the key word was "more," or if the required operation was subtraction, the key word was "less"). In inconsistent language problems, the required operation for the first step was the reverse of the operation primed by the key word (e.g., if the required operation was addition, the key words was "less," or if the required operation was subtraction, the key word was "more"). Examples of this type of problem are in Figure 1.1.

Consistent-Less

At Lucky, butter costs 65 cents per stick.
Butter at Vons costs 2 cents less per stick than butter at Lucky.
If you need to buy 4 sticks of butter,
how much will you pay at Vons?

Consistent-More

At Lucky, butter costs 65 cents per stick.
Butter at Vons costs 2 cents more per stick than butter at Lucky.
If you need to buy 4 sticks of butter,
how much will you pay at Vons?

Inconsistent-Less

At Lucky, butter costs 65 cents per stick.
This is 2 cents less per stick than butter at Vons.
If you need to buy 4 sticks of butter,
how much will you pay at Vons?

Inconsistent-More

At Lucky, butter costs 65 cents per stick.
This is 2 cents more per stick than butter at Vons.
If you need to buy 4 sticks of butter,
how much will you pay at Vons?

Figure 1.1 Consistent and inconsistent language versions of the butter problem (from Mayer and Hegarty, 1996).

The most common mistake is known as reversal error, because problem solvers perform the opposite operation of what is actually required (i.e., in the Inconsistent-more version of the butter problem in Figure 1.1, students would add 2 cents to 65 cents instead of subtracting).

Recall and Recognition of word problems

Mayer (1981) analyzed nearly 1100 algebra story problems collected from 10 standard algebra textbooks in common use in California junior high schools at the time. He identified approximately 24 families of problems based on the nature of the source formula involved (e.g., “time rate” problems were based on the formula “distance or output = rate x time”) and on the general form of the story line (see Mayer, 1981, 1982). Each family was divided into templates based on the specific propositional structure of the problem, yielding a total of approximately 100 templates or problem types. Some problem types were very rare, occurring only once or twice out of 1100 problems (which actually means the problem only appeared at all in one or two of the textbooks). Other problem types were much more common, occurring anywhere from 9 to 40 times per 1100 problems.

When college students were asked to recall a series of eight algebra story problems, Mayer (1982) found that relational statements were approximately three times more likely to be mis-recalled than assignment statements, and that problem types commonly found in mathematics textbooks were more easily recalled than uncommon problem types. Students were much more likely to mis-recall relational statements as assignment statements than to mis-recall assignment statements as relational statements, and although students sometimes converted a less common problem type into a more common one, the reverse never occurred. Cummins, Kintsch, Reisser, and Weimer (1988) also found that students tended to miscomprehend difficult word problems by converting them into simpler problems.

Hegarty, Mayer and Monk (1995) asked college students to solve a series of 12 word problems which included 4 target two-step problems with relational statements described earlier. Students were then asked to recall the 4 target problems (by asking them to write down the problem about “butter,” etc.), and to take a recognition test where they were asked to identify which of the four possible forms (see Figure 1.1) each of the four target problems had taken.

Unsuccessful problem solvers were more likely to make semantic errors (i.e., to remember the exact wording of the relational key word but not the actual relation between the variables) in recalling and recognizing problems than successful problem solvers, who were more likely to make literal errors (i.e., to remember the correct relationship between the variables but not the actual key word, thus retaining the correct meaning of the problem), such as in Figure 1.2.

Original Problem

At Lucky, butter costs 65 cents per stick.
This is 2 cents less per stick than butter at Vons.
If you need to buy 4 sticks of butter,
how much will you pay at Vons?

Semantic Error

At Lucky, butter costs 65 cents per stick.
Butter at Vons costs 2 cents less per stick than butter at Lucky.
If you need to buy 4 sticks of butter,
how much will you pay at Vons?

Literal Error

At Lucky, butter costs 65 cents per stick.
Butter at Vons costs 2 cents more per stick than butter at Lucky.
If you need to buy 4 sticks of butter
how much will you pay at Vons?

Figure 1.2 Semantic and literal errors in remembering the butter problem, (Mayer and Hegarty, 1996).

Learning to Solve Word Problems

A review of mathematics textbooks shows that most of the word problems can be solved by using a direct translation strategy and that in some cases, a direct translation strategy is even taught (Briars & Larkin, 1984). Lewis and Mayer (1987) examined the errors that college students made as they solved a series of word problems containing both consistent and inconsistent language problems. The overwhelming majority of errors made were reversal errors rather than computational errors. In follow-up studies, students were 5 to 10 times more likely to make reversal errors on inconsistent language problems than on consistent language problems (Hegarty, Mayer & Green, 1992; Lewis, 1989).

College students who showed a pattern of making reversal errors on inconsistent but not consistent problems were given two sessions of instruction on how to represent word problems within the context of a number line diagram (Lewis, 1989). Students who received representational training showed large pretest-to-posttest reductions in problem solving errors on word problems, whereas the control group of students who did not receive the training did not show large reductions. Problem solving errors were virtually eliminated in the group of students who received representational training, whereas error rates in the control group, which did not receive the training, remained unchanged.

In one strand of research, Hegarty and Mayer and their associates (Hegarty, Mayer & Green, 1992, Hegarty, Mayer & Monk, 1995) examined the eye fixations of high school and college students as they read a series of word problems presented on a computer monitor. The student's task was to describe how they would solve the problem (Hegarty, Mayer and associates used a fixed-head eye-tracking system, thus students were unable to make written calculations to actually solve the problems). The target two-step consistent and inconsistent language problems were presented among a variety of one- and two-step problems. Successful problem solvers were defined as those students who made no more than one error in planning solutions to the problems. Hegarty, Mayer and Green (1992) defined unsuccessful problem solvers as those students making two or more errors, which seems rather arbitrary, but Hegarty, Mayer and Monk (1995) replicated their results using a more conservative criterion of four or more errors. The most common error was a reversal error.

Successful problem solvers spent more time reading inconsistently worded problems than they did reading consistently worded problems. They spent that extra time by rereading variable names more in inconsistent problems than in consistent problems. Successful problem solvers spent more time on inconsistent than on consistent problem while less successful problem solvers spent about the same amount of time on both types of problems. Less successful problem solvers focused a larger proportion of their rereading on numbers than did successful students who focused a larger proportion of their rereading on variable names.

Mayer and Hegarty argue that their research provides converging evidence that students often emerge from K-12 mathematics education with adequate problem execution skills, but inadequate problem representation skills. The pattern of reversal errors on inconsistent but not consistent problems seems to support the idea that unsuccessful problem solvers use a direct translation strategy. They conclude that the source of difficulty in mathematical problem solving is in problem representation rather than solution execution. Furthermore, the source of difficulty in problem representation is in comprehension of relational statements rather than assignment statements and the source of difficulty in understanding relational statements involves using a direct translation strategy rather than a problem model strategy.

Students who use a key word approach see the word "less” and are inclined to subtract. Briars and Larkin (1984) have shown that a key word approach to understanding word problems can be effective for many problems commonly found in mathematics textbooks. The difference between a key word approach and a model construction approach to problem representation may exemplify a possible difference between successful and unsuccessful problem solvers.

Finding a balance between simplified problems that assist the learner to discover the semantic structure underlying the problems on the one hand, and presenting a broad selection of problems to challenge children to think rather than to just do, will be a complicated process, in part because children vary a great deal in their learning styles and what is right for one child will not work for another. Something to keep in mind throughout all of this is that while average US children lag behind their age-mates in many other countries on international assessments of mathematics achievement, the standard deviations are quite large and the best US children are on par with the best in the world. This is not good enough however, because as Stevenson, Lummis, Lee and Stigler (1990) discovered, although US fifth graders who are matched on tests of computational ability outperform their Japanese counterparts on tests of problem solving ability, only 5% of the US children tested performed at the highest level compared to 77% of the Japanese students tested. Which is to say, the best of the American children (95th percentile in the language of standardized US tests) are on par with the majority of the Japanese children tested (those in the 33rd percentile). This should be quite sobering.

It is also cause for some hope. Americans are proud of their creativity and flexibility of thinking, and the findings of Stevenson et al. (1990) reflect that. Japanese teachers express concern that the emphasis on entrance exams (to university, to competitive high schools, and in some cases to junior high or even elementary schools) and the tendency to teach for those exams reduces the creativity their students bring to the problem solving process. In the process, however, schools in the US seem to fail to address the needs of the majority, as evidenced by the poor performance of US children compared to their counterparts in other countries.

The difficulty with word problems may also arise from the predominance of canonical representations of problems. When children are learning their basic mathematical facts – addition, and multiplication of single digit combinations, and the equivalent subtraction and division combinations – there is an overwhelming tendency to present canonical presentations (e.g., $2 + 3 = ?$ rather than $2 + ? = 5$ or $? + 3 = 5$ or $? = 2 + 3$ or any of the other six left-handed variations on this number sentence) of problems. Thus when children attempt to translate a word problem into a numerical sentence, they are less familiar with non-canonical representations (which may in fact be a more natural way of modeling the problem). One wonders if children who are taught a variety of ways of representing simple addition and subtraction problems may find it easier to solve word problems, because they will have a more flexible method of representing the syntactic structure of problems. For example, children are not as successful on Change 3 problems (e.g., Pete had 3 apples; Ann gave him some more apples; now Pete has 10 apples; how many apples did Ann give to Pete?), Can children successfully represent this as $3 + ? = 10$, and does that representation assist with solution?

CHAPTER TWO:
SOLVING ARITHMETIC WORD PROBLEMS

Experiments

Stevenson, Lee and Stigler (1986) have demonstrated that the disparity in mathematics achievement between Asian and American children exists as early as the first grade. Miura (1987) suggests that these differences may be the result of differences in cognitive understanding of number resulting from the relative ease with which the Base 10 numeration system maps onto the number words of languages based on ancient Chinese. She has documented differences in understanding of the concept of place value in Asian speaking first-graders compared to their American counterparts. Differences in mathematical achievement in younger children across cultures other than Asian countries, however, have remained largely untested. Both the first and the second international assessments of mathematical achievement have focused on the achievement of older children, 13- and 17-year-olds, by which time socio-cultural factors such as schooling may have had significant influence.

If the performance advantage experienced by Asian children is linguistic as Miura and her colleagues (Miura, Kim, Chang & Okamoto, 1988; Miura, Okamoto, Kim, Steere & Fayol, 1993) suggest, then one would not predict differences in mathematics achievement between first-grade students in various European countries whose languages do not reflect the regularities of the Base 10 numeration system. First-grade students in France, for example, who perform similarly to American first-graders on tests of place value skills should perform similarly on tests of mathematical achievement but to my knowledge, there are no studies comparing the mathematical achievement of French and American children in primary school.

Although it is still unclear how early the differences in mathematical achievement become evident, by age 13, U. S. children clearly lag behind their counterparts in many countries. Stigler et. al. (1986) noted that US textbooks present a preponderance of the types of word problems that American children find the easiest. It seems just as logical to conclude that American children find these types of problems easier because they have had a lot of experience solving them. The Soviet textbook series was found to present roughly equal numbers of the different types of addition and subtraction word problems, but there does not appear to be any data on how well Soviet children solve various kinds of word problems. With the dissolution of the Soviet Union, such a comparison may be much more difficult to do, but if there are still countries using the Soviet textbook series, it would be interesting to see if children using this textbook series show improved performance on types of problems that occur infrequently in American textbooks.

There has been some success teaching children how to deal with word problems they don't generally do well on. It is not difficult to conclude that if children are only presented with simple forms of problems that can be solved by directly extracting a solution procedure from the text, that they will learn to look for key words and thus answer such problems more rapidly. After all, such a heuristic has worked correctly on most of the problems they have practiced on and there is no reason for them to abstract a more general understanding of the problem in order to solve it.

Experiment 1: Replication/Verification of Problem Type Difficulty

Experiment 1 investigates first- and second-grade students' ability to solve simple, one-step addition and subtraction word problems. The purpose of this experiment is to replicate previous work on the relative difficulty of various types of simple arithmetic word problems. It is intended to verify that the problems selected have the same pattern of relative difficulty of solution as has been previously documented (Carpenter & Moser, 1982; Fuson, 1992a, b) and also to establish a baseline for subsequent comparisons.

Participants

Eighty-seven first- and second-grade students participated in these studies. There were 44 first-grade children (15 boys and 29 girls) and 43 second-grade children (16 boys and 27 girls). Mean ages at the two grade levels were 6 years-5 months for first-grade (range: 5-1 to 6-10) and 7-5 for second-grade (range: 5-11 to 8-0). Two of the second-grade children, both girls, did not report birth dates. These students were permitted to participate in the study but were obviously not included in the calculation of ages. The mean age of the second-grade students therefore reflects that of 41 children (25 girls) rather than 43. Students attended a public elementary school in a major East Coast city serving a middle to lower income neighborhood and were predominantly African American. Two first-grade students failed to take one of the tests due to experimenter error and were not included in analyses including that test.

Materials

Materials consisted of two sets of written word problems. There was a set of 6 typically worded Change problems, and a set of 5 Compare problems.

One of the Compare problems was included for a different comparison and will be discussed later. Each set of problems was a single page long. Problems were typed in a moderately large, easy to read font and single spaced with a substantial amount of white space between each problem. All numerals presented in the problems were written as Arabic numbers. Examples of these problem sets may be found in Appendices A and C.

The set of Change problems included a single example of each of the six different types of Change problems. There was an addition example and a subtraction example each of the three different types of change problems: solve-for-result (Change 1 and 2), solve-for-change-set (Change 3 and 4), and solve-for-start-set (Change 5 and 6).

The four Compare problems with which this study is concerned included a pair of problem in which the comparison cued the correct operation (Compare 3 and 4) and a pair of problems in which the comparison cued the opposite operation (Compare 5 and 6). Each of these pairs included an addition problem and a subtraction problem. Compare 1 and 2 problems were not included due to concerns about timing since piloting indicated that it took considerably longer for children to finish the Compare set than it did for them to complete the Change set.

In both sets of problems, if there were two actors involved, one actor was male and the other was female. There were two different problem orders for the set of Change problems and five different problem orders for the set of Compare problems. Students were randomly assigned to one of the orders for each set of problems. An independent-samples t-test of the Change problems and an ANOVA for the Compare problems revealed no significant differences due to problem order. The data were therefore collapsed across order.

Procedure

Each student was tested individually. Most students were pulled out of class and taken to a quiet area to work. This was generally at a table in a quiet hallway but occasionally students worked in the 'library' corner of their own classroom or in an empty classroom if one was available. First-grade students were tested in 2 sessions, generally on different days and usually a few days, but no more than a week, apart. Each session consisted of a single set (page) of problems. Second-grade students were generally tested in a single session. Occasionally a second-grade session was interrupted but there is no evidence that these interruptions affected student performance. In these cases, the test was completed at the next opportunity - the same day if possible. Both first- and second-grade sessions generally lasted about 15-20 minutes. Students were audio-taped as they solved the set of word problems and questioned about some of the problems after they completed the problem set. Students were randomly assigned to receiving the Change problems first or the Compare problems first.

Students were asked to follow along as the researcher read each problem out loud. "We're going to do some math problems. These are story problems so you'll have to figure out how to find the answer. Some of them are addition problems and some of them are subtraction problems but I can't tell you which ones are which. You have to figure it out. I'll read the problem out loud for you and then you can solve it. I'll read the problem as many times as you want me to. You can use anything you want to help you figure out the answer. You can use the things in my pencil box or your fingers or you can even use my fingers. You can also make marks on this blank piece of paper. When you are all done, I'm going to ask you some questions about what you did. Just because I ask you questions doesn't mean your answer is wrong, OK?"

After receiving a clear acknowledgement that the student understood, the experimenter asked the student to answer a simple arithmetic problem such as "What's one plus one?" After the student responded, she was asked to explain "How do you know that?" The experimenter probed the student two or three times as needed to encourage elaboration beyond ambiguous answers such as "because…". This school encouraged students to explain their answers as part of the curriculum and none of the students attempted to change their answers when asked about how they figured that out.

Each problem was read out loud at least twice. The researcher asked the student if he would like to hear it again and if the student seemed reluctant the researcher reminded the student that she would read the problem as many times as he wanted. Students were given as long as they needed to solve the problem. The use of manipulatives to assist with counting was encouraged. Students were told that they could use the contents of the researcher's pencil box, their own fingers or the fingers of the researcher. The researcher's pencil box contained several writing instruments, an eraser, some paperclips and a spare pair of double A batteries for the micro-cassette recorder, intentionally enough items to solve any of the problems presented. Students were also given a blank sheet of paper and told that marks could be made upon on it to help them solve the problems. Most students used some combination of these methods. If a student chose to use the researcher's fingers, he was told that he had to manipulate the fingers up and down himself. After the student completed the problem and wrote down the answer, he was asked to "write down the math problem you used to figure out the answer to my question." Once the student completed the problem, he was asked if he was ready to continue and the next problem was read aloud.

Once the student completed all of the problems being presented in the session, the researcher asked if she could ask him some questions. She would then point to a problem and ask the student to explain how he got his answer. The researcher would probe gently for expansion and clarification and then thank the student for his answer. In general, the researcher would ask first-grade students about approximately half of the five or six problems in the session. Second-grade students were usually asked for explanations about more problems since second-grade sessions included more problems than a first-grade session. The researcher tried to include at least one problem that the student answered correctly when asking for explanations.

Scoring

A student's answer on a given problem was scored as correct if they wrote down the correct numerical answer. In order to avoid experimenter bias, no allowances were made for counting errors observed by the experimenter (such as the child double counting an object) or for recording errors such as the juxtaposition of digits (e.g., child says 16 and writes "61").

Results

Although boys appear to be slightly better (mean 4.7 problems correct) than girls (mean 4.0 problems correct) at solving these problems overall, this difference is not significant ($p=.149$). The difference between boys and girls on subtraction problems is also not significant ($p=.116$). There is no difference between boys and girls on addition problems overall. There is also no difference in the relative difficulty of solution for addition and subtraction problems either overall (p=.110) or by gender. If anything, the raw scores suggest that children, boys in particular, may find these subtraction problems slightly easier to solve than the addition problems though there is statistically no significant difference in their performance (see Table 2.1). This finding is somewhat surprising as subtraction problems are generally considered to be more difficult for children to solve than addition problems.

Table 2.1 Percent correct overall by arithmetic operation and gender.

	Boys (N=29)	Girls (N=56)	Overall (N=85)
Addition	.4414	.3893	.4071
Subtraction	.5034	.4143	.4447
Total	.4724	.4018	.4259

It is not really surprising to find that second-grade children are better overall at word problem solution than first-grade children ($p<.001$), given their additional year of mathematics instruction and practice. Second-grade children correctly solved an average of 5.3 of the ten problems where first-grade children correctly solved an average of 3.1 of the same ten problems. Second-grade children do better than first-grade children on both addition and subtraction problems ($p<.001$, both).

Table 2.2 Addition and subtraction problems, percent correct by grade

Grade		Mean % correct	Std. Dev.	Minimum # correct	Maximum # correct
1	Addition	0.3143	0.1539	0	3
N=42	Subtraction	0.3286	0.2156	0	4
	Overall	0.3214	0.1507	0	7
2	Addition	0.4977	0.2559	0	5
N=43	Subtraction	0.5581	0.2249	1	5
	Overall	0.5279	0.2175	2	10

Informal conversations with teachers indicated that the first-grade students had not yet been introduced to subtraction at the time these tests were administered, but this is not evident from the student scores. There is no significant difference between first-grade children's performance on the addition problems and the subtraction problems. Second-grade children, however, appear to do slightly better on the subtraction problems than they do on the addition problems (p=.062). There are no gender differences in performance overall at either grade level (see Table 2.3) even though second-grade girls are better at solving subtraction problems than addition problems (p=.039). These girls solved 54.07% of the subtraction problems correctly and 46.67% of the addition problems correctly. Second-grade boys do not demonstrate this disparity between addition and subtraction problems. Nor do first-grade students of either gender although first-grade boys show a trend to score better on subtraction problems (40.0% correct) than on addition problems (30.77% correct) that may be marginally significant (p=.111).

Table 2.3 Mean percent correct by grade and gender.

Grade	Gender	N	Mean	Std. Dev.	Std. Error Mean
1	Male	13	0.3538	0.1808	0.0502
	Female	29	0.3069	0.1361	0.0253
	Total	42	0.3214	0.1507	0.0233
2	Male	16	0.5688	0.2626	0.0657
	Female	27	0.5037	0.1870	0.0360
	Total	43	0.5279	0.2175	0.0332

Change problems are broadly reported to be easier for children to solve than Compare problems. The sets of problems used in this experiment replicate these findings. Children are much more successful solving Change problems than they are at solving Compare problems. This is true for both boys and girls and at both grade levels at p<.001 (in all four cases) and for boys and girls within each grade level (p=.01 or below). Boys and girls perform comparably to one another on both Change problems and Compare problems. Performance between addition and subtraction Change problems is comparable except that second-grade girls show a trend towards performing better on subtraction problems (64.20% correct) than they do on addition problems (55.56% correct) which is marginally significant (p=.070). First-grade boys also show a trend better performance on subtraction problems but on Compare problems rather than Change problems (p=.104). They get more than twice as many subtraction problems correct (26.92%) than they do addition problems (11.54%). There were only 13 first-grade boys in the study, however, so this results needs to be weighted appropriately.

As we would expect, second-grade students correctly solved significantly more problems than first-grade students on both the Change and the Compare sub-tests. First-grade children solved 41.67% of the Change problems correctly and 17.26% of the Compare problems correctly. Second-grade children solved 64.34% of the Change problems and 35.47% of the Compare problems correctly.

Table 2.4 Mean percent correct for Change and Compare problems by grade

		Change Problems			Compare Problems			Overall		
Grade		N	Mean # (%) correct	Std. Dev.	N	Mean # (%) correct	Std. Dev.	N	Mean # (%) correct	Std. Dev.
1	Boys	15	0.4444	0.2648	13	0.1923	0.2317	13	0.3538	0.1808
	Girls	29	0.4023	0.2294	29	0.1638	0.2244	29	0.3069	0.1361
	Total	44	0.4167	0.2398	42	0.1726	0.2242	42	0.3214	0.1507
2	Boys	16	0.7188	0.2836	16	0.3438	0.3521	16	0.5688	0.2626
	Girls	27	0.5988	0.1806	27	0.3611	0.3203	27	0.5037	0.1870
	Total	43	0.6434	0.2288	43	0.3547	0.3284	43	0.5279	0.2175

Results by problem

Overall, solve-for-result (Change 1 and 2) problems are easier for students to solve than solve-for-change-set (Change 3 and 4) problems ($p<.001$), which are in turn easier to solve than solve-for-start-set (Change 5 and 6) problems ($p<.001$). This is as expected and true for both boys and girls and at both grade levels and is significant at $p=.02$ or better. First-grade boys do not perform significantly better on solve-for-change-set problems than they do on solve-for-start-set problems ($p=.301$) though second-grade boys do ($p=.027$). These findings regarding the relative difficulty of different sorts of Change problems broadly replicates those reported in the literature (Carpenter & Moser, 1982; Riley et al., 1983; Fuson, 1992 a, b).

The six different kinds of Change problems can be ranked according to difficulty based on the percentage of children solving each problem correctly. Overall, from least difficult to most difficult, these Change problems would be Change 1 (92% correct), Change 2 (71%), Change 4 (61%), Change 3 (46%), Change 6 (33%) and Change 5 (14%). Boys and girls both show this same pattern of success rate with no significant differences between the percentage of boys and girls answering correctly except on the Change 5 problem where the boys meet with greater success solving the problem correctly ($p=.002$). 29% of the boys answered this problem correctly and only 5% of the girls answered this problem correctly. A closer look indicates that this result is due to the performance of second-grade students. This problem was correctly answered by 44% of the second-grade boys and only 7% of the second-grade girls (p=.002). The difference in correct response rate between first-grade boys (17%) and first-grade girls (3%) is not significant.

Table 2.5 Percentage of students answering correctly on Change problems

	Solve for result		Solve for change set		Solve for start set	
Grade	Change 1	Change 2	Change 3	Change 4	Change 5	Change 6
1	.89	.55	.27	.45	.07	.27
2	.95	.88	.65	.77	.21	.40
Total	.92	.71	.46	.61	.14	.33

First- and second-grade students have nearly the same pattern of successful responses as is seen overall although first-grade students show no difference in their ability to solve Change 3 and Change 6 problems (see Table 2.5). Second-grade students perform better than first-grade students at most of the problems. Second-grade students outperform first-grade student on Change 2, 3 and 4 problems, and show a trend in that direction (p=.057) on Change 5 problems. There is no difference between first- and second-grade students' performance on Change 1 problems or Change 6 problems. Students in both grades do well on Change 1 problems: 89% of the first-grade students and 95% of the second-grade students solve this problem correctly. Students in both grades perform relatively poorly on Change 6 problems: only 27% of first-graders and 40% of second-graders answer this problem correctly.

The results are somewhat less definitive for Compare problems. Children perform more poorly on Compare problems that they do on Change problems but contrary to suggestions in the literature (Carpenter & Moser, 1982; Fuson, 1992, a, b) that Compare problems as a class are more difficult than any type of Change problem, the Compare problems tested seem to be comparable in difficulty to the more difficult Change problems but not necessarily more so.

Experiment 2: Typical vs Clarified Wording (Change problems)

Hudson (1983) demonstrated that children as young as kindergarten perform better on Compare problems which have been formulated to take advantage of children's ability to use a matching strategy to make a correspondence between items in the problem more obvious versus more standard forms of the question. Hudson's data suggest that children's difficulties on Compare problems are influenced by the formulation of the problem.

Based on these findings, DeCorte, Verschaffel and DeWin (1985) systematically tested the hypothesis that rewording simple addition and subtraction word problems to make the semantic relations are made more explicit would facilitate the solution of these problems by young elementary school children. A group of first- and second-grade students were tested on two sets of Combine 2, Change 5, and Compare 1 problems. In one set, the problems were reformulated to make the semantic relations more explicit and the other set was left in the usual 'condensed' form. The reworded problems were solved significantly better than the standard problems. DeCorte and Verschaffel (1987) hypothesize that less able and inexperienced children depend more on text driven processing to construct an appropriate problem representation because their semantic schemes are not very well developed.

While rewording problems to make the semantic structure more explicit may assist younger children in solving a greater range of problems, we must not forget that in the long run, children must learn to solve more tersely worded problems and problems with more complicated wording.

The purpose of this experiment was to determine whether relatively simple or minor clarification of semantic and temporal relationships has an effect on the solution of Change problems. That is, children were tested to see whether simple language clarification is able to improve their rate of successful problem solution.

Participants

The same 87 first- and second-grade children participated in this study as participated in Study 1. One first-grade student was eliminated from the analyses because student did not take one of the two tests resulting in 43 first-grade students rather than 44.

Materials

Materials consisted of a set of 6 typically worded Change problems and a set of 6 Change problems in which the semantic relationships were clarified. The typically worded problems were the same as those used in Experiment 1 (see Appendix A). In the set of clarified problems (see Appendix B), all pronouns were replaced with personal pronouns to reduce ambiguity of reference. Verb tense changed during the problem to reflect the passage of time and to clarify action. The initial sentence of the problem was changed to the past tense if it was not already past tense. The final question was asked in the present tense. Temporal cues such as “then” and “now” were added to emphasize the temporal order of action and to emphasize temporal cues. In both sets of problems, if there were two actors involved, one actor was male and the other was female (see Figure 2.1). When constructing these problems, an effort was made to avoid using gender ambiguous names such as Sandy, Toni or Robin as an additional aid for clarification.

Typical Change 2 problem:

David had 11 cookies.
He gave 4 cookies to Sharon.
How many cookies does David have now?

Clarified Change 2 problem:

Nancy had 6 brownies.
Then *Nancy* gave Oliver 4 brownies.
How many brownies does Nancy have now?

Figure 2.1 Examples of Typical and Clarified Change 2 problems

There were 2 different orders for each set of problems. Independent-samples t-tests for both the Typical problems and the Clarified problems indicate that there was no effect of order, therefore the data were collapsed across order.

Procedure

The procedure and instructions were identical to Study 1. Students were pulled out of class and tested individually. First-grade children were tested in 2 different sessions; second-grade students were tested in a single session whenever possible. The problems were read aloud to students at least twice and as many times as the student wanted. After students completed the set of problems, they were questioned about how they got their answers for some of the problems. The two problem sets (Typical and Clarified) were administered in a random order.

Results

As with Experiment 1, boys (mean 7.0/12 problems correct) do not differ significantly from girls (mean 6.2/12 problems correct) at solving these problems overall. This is true for both addition problems and subtraction problems (see Table 2.6). Subtraction problems are actually solved better than addition problems overall ($p=.002$). This finding is counterintuitive as subtraction problems are generally considered to be more difficult for children to solve than addition problems. Although both boys and girls appear to do somewhat better on subtraction problems than addition problems, only the results for girls are significant ($p=.001$). Exploring this further we find that this gender disparity continues (see Table 2.7). Both first- and second-grade girls are significantly better at subtraction than they are at addition (1st: $p=.012$; 2nd: $p=.025$) whereas boys at both grade levels perform comparably on addition and subtraction problems.

Table 2.6 Percentage and number of problems correct by computation type and gender

	Boys (N=31)	Girls (N=55)	Overall (N=86)
Addition	.5645 (3.4)	.4667 (2.8)	.5019 (3.0)
Subtraction	.6022 (3.6)	.5636 (3.4)	.5775 (3.5)
Total	.5833 (7.0)	.5152 (6.2)	.5397 (6.5)

The results of this experiment are somewhat disappointing. Overall, performance on Typical and Clarified Change problems is comparable. No significant differences were found between them. This is true for both grades and no effects of gender were found at either grade or overall. None of the first-grade children got all of the problems correct (maximum score: 11/12) and none of the second-grade children got all of the problems incorrect (minimum: 2; maximum: 12).

Table 2.7 Percentage and number of addition and subtraction problems answered correctly by boys and girls

Grade	Gender	N	Addition Percent (#) Correct	Subtraction Percent (#) Correct	Overall Percent (#) Correct
1	Boys	15	.4222 (2.5)	.4222 (2.5)	.4222 (5.1)
	Girls	28	.3631 (2.2)	.4762 (2.9)	.4196 (5.1)
	Total	43	.3837 (2.3)	.4574 (2.7)	.4205 (5.0)
2	Boys	16	.6979 (4.2)	.7708 (4.6)	.7344 (8.8)
	Girls	27	.5741 (3.4)	.6543 (3.9)	.6142 (7.4)
	Total	43	.6202 (3.7)	.6977 (4.2)	.6589 (7.9)

Once again we find that second-grade children are better than first-grade children at both types of problems.

What is most striking about the Clarified problems is that subtraction problems are solved significantly better than addition problems ($p=.004$) overall. Boys do not show this differential performance but girls do ($p=.001$), in particular, first-grade girls ($p=.016$). Second-grade girls show a trend in this direction which is marginally significant ($p=.096$). There are nearly twice as many girls as boys in this study so these results may be somewhat more reliable.

Table 2.8 Mean percent correct for Typical and Clarified problems by grade

		Typical Problems			Clarified Problem			Overall		
Grade	Gender	N	Mean % Correct	Std. Dev.	N	Mean % Correct	Std. Dev.	N	Mean % Correct	Std. Dev.
1	Boys	15	0.4444	0.2648	15	0.4000	.3321	15	.4222	.2772
	Girls	29	0.4023	0.2294	28	0.4286	.2461	28	.4196	.2097
	Total	43	0.4167	0.2398	43	0.4186	.2755	43	.4205	.2321
2	Boys	16	0.7188	0.2836	16	0.7500	.2981	16	.7344	.2809
	Girls	27	0.5988	0.1806	27	0.6296	.2972	27	.6142	.2194
	Total	43	0.6434	0.2288	43	0.6744	.2998	43	.6589	.2479

The most obvious reason that differences were not found is that it is possible that there was insufficient difference between the problem types for children with little experience at formal arithmetic and minimal experience with arithmetic word problems. The clarifications made were in fact pretty subtle, especially when one takes into consideration certain conventions that were followed for both sets of problems, such as attempting to eliminate gender ambiguous names from problems and using actors of different genders in problems with more than one actor. The latter may have unintentionally resulted in a tendency to select problems which had two actors and to reduce the number of pronouns used in the typical problems. Choices such as these may have unintentionally contributed to a greater similarity between clarified and typical problems than was intended. Another possibility is that the subtraction effect was an artifact of the problems chosen, something that cannot be discounted given the restricted set of problems. Mautone (1999) found marginally significant differences in performance found when temporal and spatial modifiers were added to Change 1 and 2 problems appeared to be confounded by verb choice. She - speculated that language plays a complex role in children's understanding of word problems and suggested that effects of language need to studied more systematically. This suggestion is concordant with Carpenter (1985) who suggested that subtle differences such as verb tense could have an effect of children's solution of word problems.

Results by problem

As with Typical problems, solve-for-result problems are correctly solved more often than solve-for-change-set problems ($p<.001$). Both boys ($p=.003$) and girls and both grade levels at $p<.01$ or better. Girls at both grade levels ($p=.000$, both) as well as first-grade boys ($p=.015$) show this pattern. Second-grade boys demonstrate a trend in this direction that is marginally significant ($p=.083$). There is no overall difference, however, between the correct solution rate of solve-for-change-set and solve-for-start-set problems. First-grade boys are the only students better at solve-for-change-set problems than solve-for-start-set problems ($p=.019$). Neither their female classmates nor second-graders of either gender solve Change 3 and Change 4 problems better than Change 5 and Change 6 problems.

Ranking the six Clarified change problems in order of increasing difficulty is somewhat less clear than it was for Typical change problems since somewhat different patterns of success are seen both by grade and by gender. Overall, from least to most difficult, these clarified problems would be ranked: Change 1 (78% correct), Change 2 (73%), Change 4 (53%), Change 6 (52%), Change 3 (37%) and Change 5 (34%). Unlike with Typical problems, the percentage of students solving Change 1 and Change 2 problems correctly is not significantly different. Change 4 and 6 problems correctly are also virtually indistinguishable from one another as are Change 3 and 5 problems.

Other than a trend for boys to answer correctly more frequently than girls on Change 3 problems that is marginally significant (p=.110) boys and girls perform similarly to one another. First-grade boys show a marginal trend in this direction that is also not significant (p=.150). Curiously, first-grade girls do much better than first-grade boys (p=.010) on Change 6 problems. 61% of girls get this problem correct where only 20% of boys do.

Table 2.9 Percentage of students answering correctly on Clarified change problems

	Solve for result		Solve for change set		Solve for start set	
Grade	Change 1	Change 2	Change 3	Change 4	Change 5	Change 6
1	.70	.58	.21	.40	.16	.47
2	.86	.88	.53	.67	.51	.58
Total	.78	.73	.37	.53	.34	.52

Second-grade students outperform first-grade students except on Change 6 problems and Change 1 problems. On the latter, they show a trend in that direction (p=.070).

Comparing the percentage of students answering correctly on Typical and Clarified problems suggests that students are not as successful at solving Clarified solve-for-result problems as they are at solving Typically worded problems. It also suggests that students experience greater success at solving Clarified solve-for-start-set problems. These finding are borne out by paired-samples t-tests. Although there are no overall differences found between Typical and Clarified problems on solve-for-change-set problems, second-grade girls are significantly better at Typical solve-for-change-set problems than they are at Clarified ones. This results in a marginally significant trend by grade (p=.110) and a suggestive trend by gender which is not significant (p=.151). There is an overall difference evident on solve-for-start-set problems. Clarified problems are solved with greater success than Typical problems (p=.021). First-grade girls improve more on Clarified change-for-start-set problems. This is significant at p=.054. They actually do not improve on Typical solve-for-start-set problems at all. There is a marginally significant trend in the same direction by second-grade girls (p=.081). There are no differences found between Typical and Clarified solve-for-result problems.

Table 2.10 Percentage of students answering correctly on Typical and Clarified problems

		Solve for result		Solve for change set		Solve for start set	
Grade	Type	Change 1	Change 2	Change 3	Change 4	Change 5	Change 6
	Typical	.89	.55	.27	.45	.07	.27
1	Clarified	.70	.58	.21	.40	.16	.47
	Typical	.95	.88	.65	.77	.21	.40
2	Clarified	.86	.88	.53	.67	.51	.58

The trend seen in Typical problems for greater success solving subtraction problems than addition problems on everything except solve-for-result problems is also seen in Clarified problems. More students solve Change 4 problems correctly than Change 3 problems (p=.004) and more students solve Change 6 problems correctly than Change 5 (p=.002). The results for Change 3 and 4 problems are seen among first-grade students and as a strong trend among second-grade (p=.031 and p=057 respectively). Significantly more first-grade girls solve Change 4 problems correctly than solve Change 3 problems correctly; second-grade boys also show a trend for correctly solving Change 4 problems more often which is marginally significant at $p=.083$). The results for Change 5 and 6 problems appear to be the result of first-grade girls ($p<.001$) which are strong enough to give significant the results by grade ($p=.001$) and by gender ($p=.002$).

The clarifications did not work as hoped on except on solve-for-start-set problems. On Change 1 problems, students do better on Typical problems than they do on Clarified problems ($p=.002$). First-grade girls are especially prone to this ($p=.031$) though second-grade girls show a trend in this direction that is marginally significant ($p=.103$). On Change 3 problems there is a trend for students to perform better on typically worded problems as well ($p=.103$). This trend seems to be due to girls ($p=.059$) more than boys.

More students solve the clarified version of the problem on both Change 5 ($p<.001$) and Change 6 ($p=.005$) problems. In both cases the effect is due to the performance of girls ($p=.001$ or less for both problems) though boys show a trend in that direction on Change 5 problems ($p=.083$). Second-grade girls are significantly better at the clarified problem where first-grade girls merely show a trend to do better at the clarified problem ($p=.083$). On Change 6 problems, the effect is due to the performance of first-grade girls ($p=.005$) though second-grade girls show a strong trend in the same direction ($p=.057$).

Reasons for why the clarifications made to problems had the desired effect only on solve-for-start-set problems and not solve-for-result or solve-for-change-set problems need to be explored further. It is curious that the desired effect was achieved on problems that are widely thought of as the most difficult Change problems when it actually had the opposite effect as desired on solve-for-result problems which are problems that even young students generally can solve.

Study 3: Consistent vs Inconsistent wording

This study was designed to look at whether students are better at solving problems in which the key word is consistent with the arithmetic operation required to solve the problem correctly than they are at solving problems in which the key word is inconsistent with the correct arithmetic operation. Are students better at solving problems in which the key word is consistent with the arithmetic operation necessary to solve the problem correctly than they are at problems in which the key word is inconsistent with the correct arithmetic operation for solution?

Participants

The same 87 first- and second-grade children participated in this study as participated in Study 1. Two first-grade students failed to take the Compare test and were not included in the analyses.

Materials

Materials consisted of a set of 4 Compare problems which included one of each of the following types of problems: a consistent addition problem (Compare 3), a consistent subtraction problem (Compare 4), an inconsistent addition problem (Compare 5) and an inconsistent subtraction problem (Compare 6). Consistent wording is where the key word(s) in a problem is consistent with the arithmetic operation necessary to solve the problem; inconsistent wording cues the opposite arithmetic operation. All together and more are examples of words that cue addition. Words such as less or fewer suggest that subtraction should be used to solve a problem. The less grammatically correct less was used instead of fewer because piloting suggested that students were less likely to understand the word fewer.

Consistent and inconsistent wording are really just alternate descriptions for different types of Compare problems. Examples of a consistent and inconsistent addition problems follow (see Figure 2.2). Appendix C contains an example of the one of the tests used for this experiment.

Consistent addition (Change 3) problem

Carmen caught 2 fireflies.
Jim caught 5 more fireflies than Carmen caught.
How many fireflies did Jim catch?

Inconsistent addition (Change 5) problem:

Sarah read 9 books last summer.
Sarah read 6 more books than Tim read.
How many books did Tim read last summer?

Figure 2.2 Examples of consistent and inconsistent problems

Procedure

The procedure and instructions were identical to Study 1. Students were pulled out of class and tested individually. First-grade children were tested in 2 different sessions; second-grade students were tested in a single session. The problems were read aloud to students at least twice and as many times as the student wanted. After students completed the set of problems, they were questioned about how they got their answers for some of the problems. There were 5 orders and students were randomly assigned to a problem order.

Results

Contrary to expected results, students appear to be better at solving inconsistently worded problems than they are at solving consistently worded problems. (first-grade p=0.051) Although there is no gender difference overall, first-grade girls are twice as good at solving inconsistently worded problems as they are at solving consistently worded problems (see Table 2.11). First-grade boys on the other hand show no difference at all in their ability to solve either type of problem.

Table 2.11 Mean percentage correct for consistent and inconsistent problems by grade

			Consistent		Inconsistent		Overall (Compare)	
Grade	Gender	N	Mean % Correct	SD	Mean % Correct	SD	Mean % Correct	SD
	Boys	14	0.1786	0.2486	0.1786	0.2486	0.1923	0.2317
1	Girls	29	0.1034	0.2061	0.2241	0.3158	0.1638	0.2224
	Total	43	0.1279	0.2207	0.2093	0.2934	0.1726	0.2242
	Boys	16	0.3125	0.4425	0.3750	0.3416	0.3438	0.3521
2	Girls	27	0.3704	0.4065	0.3519	0.3877	0.3611	0.3203
	Total	43	0. 3488	0.4160	0.3605	0.3672	0.3547	0.3284

Study 4: Probable and Improbable subtraction

The sequence of the numbers (and information) in the problem text also affects children's solution processes. First- and second-grade children used either adding on (when using concrete objects) or counting up (when using verbal counting strategies) from the smaller given number to solve Combine 2 problems in which the larger number was mentioned last (Verschaffel, 1984). When the larger number was given first, Carpenter and Moser (1984) reported that children in their study tended to either separate from or count down from the larger given number. These results suggest that the strategies young children use to solve addition and subtraction word problems depend not only on the semantic structure of the problem, but also on the sequence of the given numbers in the task.

This study was designed to look at the effect of an improbable subtraction task on student ability to solve a Compare problem. Young children are taught that it is “impossible” to subtract a larger number from a smaller number (e.g., to subtract 8 from 6). It is therefore conceivable that students might notice if the key words of a problem seemed to require an “impossible” subtraction and this might cause them to stop and reread the problem more carefully. Alternatively, there is evidence that an early subtraction strategy is to subtract the smaller digit or number from a larger one, regardless of how they are positioned in a problem so an impossible subtraction situation may in fact have little effect. If a subtraction problem is improbable, that is, if it appears to call for subtracting a larger number from a smaller one, is the child more likely to solve the problem correctly than if the subtraction is probable, if it calls for subtracting a smaller number from a larger one? What effect might this have on students’ solution strategies?

Three immediate possibilities came to mind. First, perhaps students would notice that the problems is suggesting an improbable computation and this will make students stop and think about the word problem further. Perhaps they would read it again more carefully and thus be more likely to parse what the problem was asking. Second, perhaps students would proceed to subtract the smaller number from the larger one without being particularly concerned about it. There is some evidence that when confronted with a multi-digit subtraction problem, novice students will subtract the smaller digit from the larger one without being concerned about which digit comes first. That is, given the problem 72 - 38 = __, a student is likely to subtract the 3 from the 7 and the 2 from the 8, yielding a final incorrect answer of 46 instead of regrouping to get the correct answer of 34. Finally, perhaps students would be stymied by the request and respond that it was not possible to solve the problem. That is, students would notice that the subtraction was “impossible” and be unable to solve the problem.

The term improbable is used rather than impossible because such a subtraction problem is not, in fact, impossible except when one is restricted to whole positive numbers. Since first- and second-grade students have not yet been introduced to negative numbers, they are effectively so restricted. For this experiment, only subtraction problems were tested since addition is commutative and it is therefore not possible to create an improbable addition occurrence.

Participants

The same 85 first- and second-grade children participated in this study as participated in Experiment 3.

Materials

The Compare test used in Experiments 1 and 3 included a fifth problem that was not used in those studies. In addition to the probable Compare 6 problems that were used in the analyses of Experiments 1 and 3, there was a second, improbable Compare 6 problem embedded within the Compare test. In the probable subtraction problem, the larger of the two numbers was presented first and in the improbable problem, the larger of the two numbers was presented last. In both cases to correctly answer the problem, one must add the two numbers together in spite of the key word “less” suggesting subtraction as the appropriate arithmetic operation. Otherwise these two problems were extremely similar in form (see examples below). It was thought that presenting the larger number last might make the subtraction seem improbable and perhaps trigger a more thoughtful response.

Probable subtraction problem

Joe missed 6 problems on the math test.
Joe missed 4 less problems than Marie missed.
How many problems on the math test did Marie miss?

Improbable subtraction problem:

Kate found 2 marbles.
Kate found 8 less marbles than Billy found.
How many marbles did Billy find?

Figure 2.3 Examples of probable and improbable subtraction problems

Procedure

The procedure and instructions were identical to Study 1. Students were pulled out of class and tested individually. Both first- and second-grade children were tested in a single session. Each problem was read aloud to the student at least twice and repeated as many times as the student wanted to hear it. After students completed the set of problems, they were questioned about how they got their answers for some of the problems. There were 5 orders and students were randomly assigned to a problem order. There were no differences found so responses were collapsed across order.

Results

There were no significant differences found between the students' performance on problems whether the subtraction was probable or improbable. At least half of the students, 26 first-grade and 22 second-grade, solved both problems incorrectly. Ten first-grade students and 13 second-grade students solved both problems correctly. Curiously enough, after eliminating those students who got either both problems incorrect or both problems correct, the remaining 14 students who solved one but not the other correctly (6 first-grade, 8 second-grade) are split exactly evenly between which problem was solved correctly (see Figure 2.4).

It may be that these problems are sufficiently difficult for children that something so minor as switching the order in which the numbers are presented has little effect on solution. Which is to say, the order in which numbers are presented in a problem may be insufficient to override the tendency to say, subtract the smaller number from the larger one. An analysis of the student responses might give insight as to what strategies students were using to solve the problem. If students are subtracting the smaller number from the larger, one would expect to see students respond "2" to the probable subtraction problem and 6 to the improbable subtraction problem.

Also, although these problems were intended to be identical except for the order of the numbers, the verb in the probable subtraction problem is "missed" while the verb in the improbable subtraction problem is "found". Missed may be considered to have a negative (subtractive) connotation where the verb in the improbable subtraction problem, found, may have a positive (additive) implication. This unintentional duplication of a subtractive term may have underscored the subtraction suggestion of the key word and rendered the probable condition more difficult to solve than the improbable condition which contains a word suggesting addition perhaps helping to counterbalance the key word suggesting subtraction. It is also possible that the additional phrase "on the math test" added to complexity of the probable subtraction problem perhaps adding to cognitive load and thus the difficulty of the problem. Both of these structural items, though unintentionally included, may have affected the relative difficulty of the probable subtraction problem.

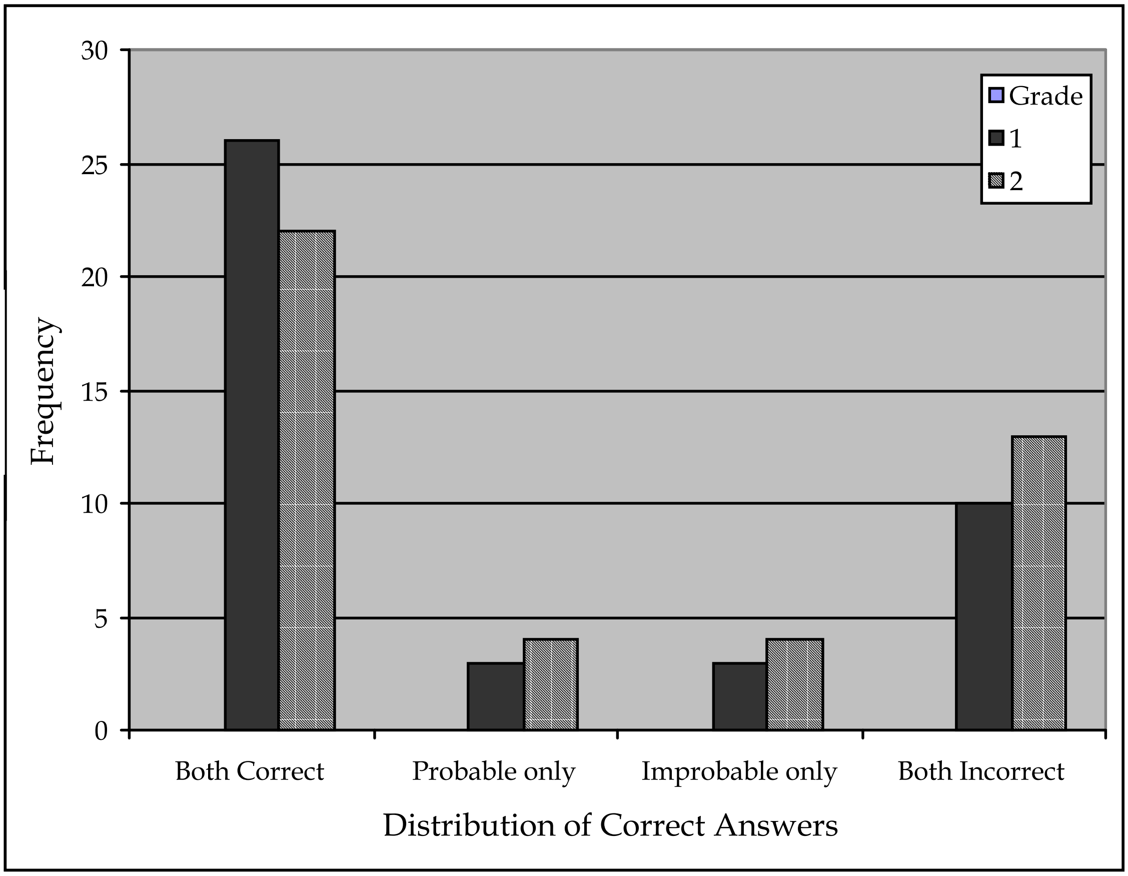

Figure 2.4 Distribution of correct answers on probable and improbable Compare 6 problems

Finally, caution must be taken in reading too much into a single pair of problems. This comparison was included because we were curious if a small change might have an effect and it appears that the change is in fact to small to have any effect. It may also be too small a change to give much idea of why not.

CHAPTER THREE:
EFFECTS OF ARITHMETIC PRACTICE ON WORD PROBLEM SOLUTION

Support for experiments

Studies of mathematics textbooks indicate that there are in fact, very few story problems included in elementary school arithmetic textbooks (Stigler, Fuson, Ham & Kim, 1986). Analyses of textbooks indicate that of the many variations of addition and subtraction word problems that are possible, the problems that children may actually encounter tend to be the simplest and easiest to solve, primarily change problems in which the result is unknown (Fuson, Stigler & Bartsch, 1986; Stigler, Fuson, Ham & Kim, 1986; DeCorte, Verschaffel, Janssens & Joillet, 1984 as reported in DeCorte & Verschaffel, 1991). By and large, however, the story problems that do appear are often presented at the end of a set of practice problems as optional problems or “challenge problems”. These problems are meant to be especially challenging for students but in practice are rarely assigned at all.

In addition, most if not all of the arithmetic practice assigned to young school children is also of the ‘solve for result’ variety. Students are asked to add two numbers together or to subtract one number from another to yield an answer. Because this form of arithmetic practice makes up the most common form of arithmetic problem to which children are exposed, these types of arithmetic problems are referred to here as canonical problems. It is important to note that the form of this arithmetic practice parallels the way one would solve a change problem in which the result was unknown.

Stigler, Fuson, Ham, & Kim (1986) conclude that textbook manufacturers are adept at giving students the kinds of problems they are good at solving. One could as easily turn that conclusion on its head and conclude that children are good at solving problems to which they are exposed and at which they receive practice solving.

If students were to get more exposure to solving for different parts of an arithmetic problem, would that exposure have an effect on their ability to solve simple (single step) arithmetic story problems?

What are canonical and non-canonical problems?

The different types of Change problems can easily be modeled as simple, single-step arithmetic sentences. Change 1 and 2 problems respectively can be modeled as addition and subtraction sentences in which the result is unknown. This is the standard form of arithmetic practice assigned to students - adding two numbers together or subtracting one number from another to determine the result (e.g., 2 + 3 = __ or 7 - 4 = __). Because this form is so common, problems in which students are asked to solve for the result after the equals sign shall be referred to as canonical problems. Canonical problems always present the problem first, before the equals sign, and expect the result to follow the equals sign.

Non-canonical problems, on the other hand, are not commonly assigned to elementary school students as practice. These atypical problems ask about one of the addends in an addition problem (e.g., 2 + __ = 5 or __ + 3 = 5), or about the subtrahend or minuend of a subtraction problem (e.g., 7 - __ = 3 or __ - 4 = 3). Change 3 and 4 problems can be modeled as arithmetic sentences in which the change set is unknown and Change 5 and 6 problems can be modeled as arithmetic sentences in which the start set is unknown. Change 3 and 4 problems are referred to by some researchers collectively as solve-for-change-set or change set unknown and Change 5 and 6 problems are sometimes called solve-for-start-set or start set unknown problems.

Finally, the presentation of problems can also be reversed. That is, rather than presenting the result last, on the right side of the equals sign, the result of the problem can be presented first, on the left side of the equals sign, with the arithmetic problem following the equals sign. (e.g., __ = 2 + 3, 5 = 2 + __ and 5 = __ + 3 or __ = 7 - 4, 3 = 7 - __ and 3 = __ -4). Since these reversed, left-handed problems are either extremely uncommon or nonexistent in arithmetic practice, they are also considered to be non-canonical problems. This includes left-handed solve-for-result (Change 1 and 2) problems.

Since Change word problems can be modeled so easily with canonical and non-canonical arithmetic word problems, it seems plausible that exposure to alternative, non-canonical, arithmetic practice may make children more aware of how to solve forms of simple arithmetic problems, namely Change problems, that they generally find difficult by giving children a larger pool of experience on which to draw to solve them. The purpose of this experiment is to see whether practice solving for numbers other than the result of adding two numbers together or subtracting one number from another has any effect on children's ability to solve simple arithmetic word problems. Specifically, if solving for one of the addends in an addition problem or the subtrahend or minuend of a subtraction problem may have an effect on a child's ability to solve Change word problems that reflect that sort of arithmetic structure.

Experiment 5: Differential arithmetic practice

If familiarity and practice play a role in children's understanding then perhaps practice at other types of problems can effect improvement in parallel arithmetic word problems. It was predicted that additional practice with non-canonical forms will improve students' ability to solve Change problems in which the change set or the start set is unknown. Although the additional arithmetic practice was not predicted to have an effect on Compare problems, these were included in the posttest for completeness.

Participants

The first- and second-grade children who participated in the earlier studies also participated in this study. Five students moved at some point during the school year and were not present to take the posttests. One student did not complete all of the worksheets and was dropped from the study. One additional student decided that he did not want to take the posttest and was excused from doing so. Of the remaining 80 children, only students who took all three of the pretests and all three of the posttests were included in these analyses. One first-grade student failed to take all three of the pretests and three first-grade students failed to take all three of the post-tests. These four students were dropped from the analyses leaving 76 students. There were 35 first-grade children (12 boys and 23 girls) and 41 second-grade children (15 boys and 26 girls). Mean ages at the two grade levels were 6 years-10 months (range: 5-11 to 7-4) and 7-11 (range: 7-5 to 8-5) for first- and second-grade respectively. The two second-grade girls for whom ages are unknown both participated in this study therefore the mean age of the second-grade children reflects that of 39 children rather than all 41 participants.

Materials

Materials for this experiment consisted of a set of arithmetic practice worksheets to be administered as class work or homework during the school year and a set of three posttests whose problems mirrored the three tests used in Experiments 1 and 2 - Typical Change, Clarified Change, and Compare - in form. There were two different sets of 44 worksheets. Each set of worksheets contained a total of 648 arithmetic problems. The majority of the worksheets contained 15 problems. The final two worksheets consisted of 14 and 4 problems respectively. Problems were either standard, canonical arithmetic practice or atypical, non-canonical arithmetic practice.

Canonical Practice

The Canonical practice consisted of problems in which two numbers were being added or subtracted and the student was asked to solve for the result. There are 81 single digit arithmetic facts when one includes all permutations of adding one through nine together. Likewise there are 81 equivalent subtraction problems with a single digit subtrahend and result. Each of these single digit addition and subtraction problems was presented a total of four times each. The worksheets were mixed addition and subtraction. Problems were randomized with the caveat that there was no problem repetition within a worksheet. Each worksheet consisted of a single page of problems typed in a large font and presented in two columns. There was sufficient space between problems for students to make tally marks or otherwise make calculations. A sample Canonical worksheet may be found in Appendix E.

Non-Canonical Practice

Non-canonical arithmetic practice problems require students to solve for one of the addends in addition problems, or for the subtrahend or minuend in subtraction problems. In addition, students were asked to solve not just "right-handed" problems in which the answer occurred at the end, after the equals sign, but also "left-handed" problems in which the answer or result was presented first and the problem followed the equals sign. This results in four possible variations/versions of each standard or Canonical problem (see Figure 3.1).

<u>Canonical problem</u>

2 + 3 = __

7 - 4 = __

<u>Non-Canonical problems</u>

2 + __= 5; __ + 3 = 5; 5 = 2 + __; 5 = __ + 3

7 - __ = 3; __ - 4 = 3; 3 = 7 - __; 3 = __ - 4

Figure 3.1 Examples of Canonical and Non-Canonical practice problems.

Each of the 81 single-digit addition problems was presented once in each configuration, requiring students to solve for one of the addends in both the right- and left-handed forms (i.e., 2 + __ = 5, __ + 3 = 5, 5 = 2 + __ and 5 = __ + 3). The equivalent subtraction problems, 5 - __ = 2, __ - 3 = 2, 2 = 5 - __ and 2 = __ - 3, were also presented once each. Each addition and subtraction problem is unique and occurs exactly once in each configuration which results in the total of 648 unique problems (see Appendix F for a sample page of problems). Problems were presented identically to those in the Canonical practice: typed, mixed addition and subtraction. It was not necessary to worry about problems repeating but for simplicity, the same arithmetic problems were used, in the same order, on equivalent Canonical and Non-canonical worksheets. Left-handed versions of solve-for-result problems (e.g., __ = 2 + 3 or __ = 7 - 4) were not included as that would have required including an addition 162 problems on both types of practice worksheets.

In both practice conditions, problems were mixed addition and subtraction, randomly ordered except for making sure all problems on a worksheet were unique. Non-Canonical and Canonical worksheets were paired and included equivalent problems in identical orders. For example, if the first problem on a Non-Canonical worksheet was 8 = __ + 5, the first problem on the equivalent Canonical worksheet was 3 + 5 = __.

Changes to Posttest

The posttest was arithmetically identical to the 3 sub-tests administered in Experiments 1 and 2. Several changes were made to the language so that the story problems were not identical but the structure of each problem was left untouched. The names of the actors in the problems were changed and the genders were switched. For example, Billy and Kate in a pretest problem became Cindy and George in the equivalent posttest problem. The object nouns used in each of the problems were also changed, generally to something similar. Thus a problem about cookies in the pretest became a problem about cupcakes in the posttest and one concerning fireflies became one about tadpoles.

Pretest

Nick gave Sue 4 marbles.
Now Sue has 7 marbles.
How many marbles did Sue have in the beginning?

Posttest

Betsy gave Rob 4 seashells.
Now Rob has 7 seashells.
How many seashells did Rob have in the beginning?

Figure 3.2 Examples of pretest and posttest Change 2 problems

Although the Compare posttest contained five problems, the improbable subtraction problem described in Experiment 4 was not used in these analyses. Thus any overall scores reflect that of 16 problems: six Typical Change problems, six Clarified Change problems, and four Compare problems.

Finally, the problems were presented in a different order on the posttest than they were on the pretest. As with the pretest there were two different orders for both types of Change problems (Typical and Clarified) and five different orders for the Compare problems. No effect of order was found so results were collapsed across order.

Procedure

Student results on the three tests used in Experiments 1 and 2 comprise the pretest or baseline for this experiment. Those studies took place during November and early December, prior to winter break. After taking the pretests, students were randomly assigned to one of three practice conditions: No Practice, Canonical Practice, or Non-Canonical Practice. Children assigned to the No Practice condition did not receive any additional practice worksheets. They were tested in the fall and in the late spring like their classmates who received practice worksheets but were not worked with in any special way during the intervening months of the school year and received no additional practice beyond the regular curriculum. Children assigned to one of the two practice conditions received a series of worksheets over the course of the spring semester. Teachers were asked to assign two to three worksheets per week as additional classroom work or as homework. Although one teacher did this successfully and two other teachers gave out some of the worksheets, the majority of teachers were unable (or unwilling) to consistently assign the worksheets. For the majority of the students, the experimenter would periodically take the children out of class to work with them in pairs or small groups on the worksheets.

The posttests were administered individually to students in the late spring (May) of the school year, approximately five months after students were initially tested on arithmetically identical word problems. The procedure paralleled that of the Experiments 1 and 2 with changes as noted. Students were pulled out of class and tested individually. First-grade children were tested in three different sessions, one session for each of the three sub-tests and second-grade students were tested in a single session. The problems were read aloud to students at least twice and as many times as the student wanted. Students were questioned about how they got their answers for some of the problems after they completed each session. For first-grade students this was at the end of each page and for second-grade students it was after they completed all three pages. The experimenter asked students about problems that they answered correctly as well as problems they answered incorrectly.

None of the students showed any inclination to change their answers when questioned about them. The order in which the three sub-tests were administered was randomized.

Although the children were randomly assigned to a worksheet condition, independent-samples t-tests were used to verify that there was no difference between the groups of students assigned to the three different worksheet conditions. Both first- and second-grade students showed an overall difference due to performance on Typical problems. First-grade students assigned to the No Practice condition did significantly better on the Typically worded pretest than their counterparts who were assigned to the Non-Canonical Practice condition ($p=.042$). Second-grade students assigned to the Non-Canonical Practice condition did significantly better than those assigned to the No Practice condition ($p=.049$). There was no significant difference between the pretest scores of those students assigned to the Canonical Practice condition and those assigned to the No Practice condition though a marginally significant trend ($p=.117$) for second-grade students in the Non-Canonical Practice condition to outperform their counterparts assigned to the No Practice condition was observed for Clarified problems. In order to accurately compare the performance of children assigned to different worksheet conditions, we must therefore look at gain scores, or the percentage of improvement rather than the percentage of problems answered correctly.

Scoring

Scoring for the posttest problems was identical to that of the pretest problems. Only correct numerical answers were scored as correct.

Overall Results

Improvement over the course of the school year is expected. Overall, single-sample t-tests of the gain scores indicate that students improve significantly at solving word problems from pretest to posttest ($p<.001$). This is true for both addition and subtraction problems ($p<.001$, both). The amount of improvement made on addition problems is not significantly different than it is for subtraction problems as indicated by a paired-samples t-test. Significant gains are made by both boys and girls ($p<.001$, both genders) at both grade levels (also $p<.001$, both grades). Both genders show improvement on addition problems and subtraction problems at $p<.001$. Although boys appear to have a slight edge over girls at solving subtraction problems (19.44% versus 13.27% improvement), this difference is not significant and there is no gender difference apparent in the amount of improvement on addition problems. The amount of improvement on addition problems is not significantly different for than for subtraction problems for either boys or girls.

Table 3.1 Percent correct and improvement by gender and overall

	Boys (N=27)			Girls (N=49)			Overall (N=76)		
	Mean	St. Dev.	Std. Error Mean	Mean	St. Dev.	Std. Error Mean	Mean	St. Dev.	Std. Error Mean
Pre	.5023	.2626	.0505	.4579	.1986	.0284	.4737	.2227	.0256
Post	.6806	.2544	.0490	.6008	.1808	.0258	.6291	.2117	.0243
Gain	.1782	.1951	.0376	.1429	.1697	.0242	.1554	.1787	.0205

Within each grade, students show the same sort of pattern of improvement that is seen overall. Both first- and second-grade students improve significantly from pretest to posttest ($p<.001$, both grades). This is true for both addition ($p<.001$, both grades) and subtraction problems (1st:

p<.001; 2nd: p=.016). There is no difference in the amount of improvement gained on addition versus subtraction problems for either grade.

Boys and girls both show improvement overall. First- and second-grade girls both show significant improvement in performance on both addition (1st: p<.001; 2nd: p=.001) and subtraction problems (1st: p<.001; 2nd: p=.050). The same is true of first-grade boys. Second-grade boys, however, show improvement only on addition problems (p=.041) and not subtraction problems (p=.178). Although the results are not significant, first-grade boys show a trend (p=.067) for greater improvement than first-grade girls. First-grade boys solve an average of five more problems correctly on the posttest than they did on the pretest for an average of 10.8 problems correct where first-grade girls averaged an increase of only three additional problems (mean 8.8 problems correct). There is no such trend evident among second-grade students. It turns out that this trend (p=.065) is due mostly to the performance of first-grade boys on subtraction problems (gain=34.38%) compared to that of first-grade girls (gain=19.57%). Although first-grade boys also appear to improve more (28.13%) than first-grade girls (17.93%) on addition problems, this difference is not significant. There are also no gender differences found among second-grade students on either addition or subtraction problems.

Table 3.2 Percentage improvement by gender and grade

	Boys			Girls			Total		
Grade	N	Mean	Std. Dev.	N	Mean	Std. Dev.	N	Mean	Std. Dev.
1	12	.3125	.1884	23	.1875	.1837	35	.2304	.1922
2	15	.0708	.1224	26	.1034	.1489	41	.0915	.1391
Overall	27	.1782	.1951	49	.1429	.1697	76	.1554	.1787

First-grade students improve significantly more than second-grade students ($p<.001$). This is true for both addition problems ($p=.014$) and subtraction problems ($p=.001$). Although first-grade boys improved more than second-grade boys ($p<.001$), the difference in improvement between first- and second-grade girls is not significant ($p=.083$). First-grade girls do not improve more than second-grade girls on addition problems ($p=.380$), but they do improve more than second-grade girls on subtraction problems ($p=.045$). First-grade boys improved more than second-grade boys on both addition and subtraction problems ($p=.002$ and $p=.003$ respectively).

If we look at performance broken out by problem class, that is, if we look at how children perform on Typical and Clarified Change problems and on Compare problems, we see that there is improvement from pretest to posttest across all three classes of problems. The mean percent correct on Typical Change problems was 53.51% on the pretest and 70.39% on the posttest. This is a gain of 16.89% ($p<.001$). Mean percent correct on Clarified change problems was 55.04% on the pretest and 71.27% on the posttest. This is a gain of 16.23% ($p<.001$). On Compare problems, there was a gain of 12.50%, with pretest and posttest averaging 26.64% and 39.14% correct respectively. Overall, neither Typical nor Clarified problems show significantly more improvement than Compare problems. This is somewhat surprising as Compare problems have widely been found to be more difficult than Change problems (Fuson, 1992a). Improvement on Typical and Clarified problems is also not significantly different from one another overall.

Boys improved on all three classes of problems (Typical: $p=.002$; Clarified: $p=.003$; Compare: $p<.001$). Girls did not improve on Compare problems ($p=.248$), but they did improve on both Typical and Clarified Change problems ($p<.001$, both). Further, although there is no difference in the amount of improvement by gender on either of the two types of Change problems (Typical or Clarified), boys show significantly more improvement on Compare problems than girls ($p=.016$). Boys get an average of 50.93% of the Compare posttest problems correct and while girls only average 32.65% of these problems correct.

Single-sample t-tests of the gain scores indicate that both first- and second-grade children improved on all three types of problems. First-grade students got an average of 2.5 of the typically worded Change problems correct on the pretest and an average of 4.3 correct on the posttest ($p<.001$). On the Clarified Change problems their pretest and posttest means scores were 2.5 and 3.9 respectively ($p<.001$). Mean percent correct on the Compare test also showed significant improvement ($p=.019$). Students scored an average of 17.86% correct on the Compare pretest and 32.14% correct on the posttest. Second-grade students averaged 4.0 problems correct on the Clarified pretest and 4.6 correct on the posttest ($p=.009$). Their improvement on Typical and Compare problems were also significant at the $p=.009$ and $p=.043$ levels respectively. First-grade students improved more than second-grade students on Typical problems ($p<.001$) and showed a trend in that direction on Clarified problems ($p=.060$). There was no difference in improvement between first- and second-grade children on Compare problems.

First-grade boys improved significantly on all three tests at $p=.01$ or better. First-grade girls also improve on Typical and Clarified problems ($p=.005$ or better). They do not however improve significantly on Compare problems. Not surprisingly, first-grade boys improve significantly more than girls on Compare problems (p=.016). Improvement in second-grade is even more clearly demarcated by gender. Second-grade boys improve significantly only on Compare problems (p=.028). Second-grade girls, on the other hand, do not show significant improvement on Compare problems but they do improve on both Typical and Clarified problems (p=.008 and p=.002 respectively). The only difference by gender among second-grade students is a trend for girls to improve more than boys on Clarified problems which is marginally significant at p=.085.

Table 3.3 Percentage of students answering correctly on pretest and posttest by problem type

Problem Type:			Typical		Clarified		Compare	
Grade	N	Gender	Pretest	Posttest	Pretest	Posttest	Pretest	Posttest
	12	Male	.4444	.7917	.3889	.6528	.2083	.5417
1	23	Female	.4130	.6812	.4420	.6449	.1630	.2065
	35	Overall	.4238	.7190	.4238	.6476	.1786	.3214
	15	Male	.7000	.7333	.7333	.7667	.3000	.4833
2	26	Female	.5890	.6667	.6154	.7692	.3654	.4327
	41	Overall	.6301	.6911	.6585	.7683	.3415	.4512

Paired-samples t-tests were used to compare the gain scores by test. First-grade students improved significantly more on Typical problems than they did on Compare problems (p=.039). Their improvement on Clarified problems is not significantly different either from that of Compare problems or Typical problems. If we break this down by gender we see that boys do not show a significant difference in the amount of improvement between any of the three tests. Girls, however, show the same pattern we see overall. First-grade girls improve significantly more on Typical Change problems then they do on Compare problems (p=.011) although their improvement on Clarified problems is not different from either Compare problems or Typical Change problems.

Second-grade students like those in first-grade also improved on all three problem types though they do not show nearly as much improvement as do first-grade students. There is, however, no significant difference in the improvement they make on the different tests. Although it appears that second-grade boys improve the most on Compare problems and second-grade girls improve more on Clarified Change problems than they do on Typical Change problems or Compare problems, these results are not significant.

Table 3.4 Percent improvement by problem type

Problem Type:			Typical		Clarified		Compare	
Grade	N	Gender	Gain	Std. Dev.	Gain	Std. Dev.	Gain	Std. Dev.
	12	Male	.3472	.2508	.2639	.2969	.3333	.3257
1	23	Female	.2681	.2343	.2029	.3096	.0435	.3167
	35	Overall	.2952	.2394	.2238	.3023	.1429	.3445
	15	Male	.0333	.1569	.0333	.1911	.1833	.2907
2	26	Female	.0769	.1352	.1538	.2207	.0673	.3575
	41	Overall	.0610	.1432	.1098	.2161	.1098	.3356

When we compare the gain scores of first- and second-grade students to one another, we find that first-grade students show more improvement than second-grade students do overall (p<.001). There is no difference in amount of improvement on Compare problems by grade. First-grade students show significantly more improvement than second-grade students on Typical

Change problems (p<.001) and show a trend in that direction on clarified Change problems though it is not significant (p=.060).

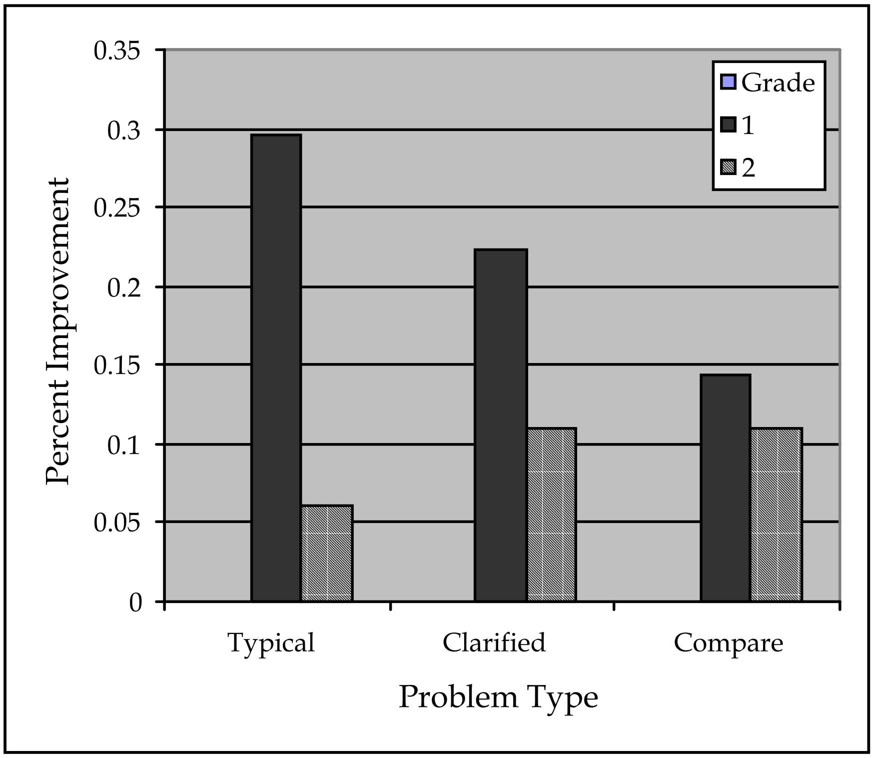

Figure 3.3 Percent improvement on Typical, Clarified and Compare problems by grade

There is a gender effect on one of the three sub-tests for first-grade children. First-grade boys do significantly better than first-grade girls on compare problems (54.1% correct versus 21.2% correct). They may in fact do better than both second-grade girls and boys on this particular sub-test Second-grade children demonstrate no effect of gender on any of the three sub-tests.

Results of Worksheet Practice

Students improved during the school year regardless of the worksheet condition to which they were assigned. Single-sample t-tests of the gain scores are significant at $p<.001$ for both students assigned to the Canonical and those assigned to the Non-Canonical Practice condition and at the $p=.001$ level for students assigned to the No Practice condition.

Children improved on both Typical and Clarified Change problems regardless of the worksheet condition to which they were assigned. There are no differences evident overall in the amount of improvement gained by worksheet condition. There are also no gender differences in the amount of improvement made.

First-grade students in all three conditions improved significantly on Typical problems. Those in the Non-Canonical and No Practice conditions also improved significantly on Clarified problems though students assigned to the Canonical Practice condition did not. Second-grade students in the Canonical Practice condition improved on both Typical and Clarified problems ($p=.014$ and $p=.026$ respectively). Second-grade students assigned to the Non-Canonical Practice condition improved only on Clarified problems ($p=.023$), not Typical problems. Second-grade students assigned to the No Practice condition did not improve significantly on either the Typical or Clarified change problems.

Although the different types of arithmetic practice were not expected to have any effect upon Compare problems, both children in the Canonical Practice condition and children in the Non-Canonical Practice condition appear to show a trend towards improvement on Compare problems ($p=.065$ and $p=.055$ respectively). Children assigned to the No Practice condition did not show the same trend of improvement on Compare problems. It appears that this trend is merely an artifact as it disappears when we examine improvement by grade level. Neither first- nor second-grade students improved significantly on Compare problems whether they were in the No Practice condition or in one of the two practice conditions.

Table 3.5 Percent improvement by worksheet condition

Grade	Worksheet Condition	N	Problem Class: Typical	Clarified	Compare
	None	14	.2143	.2143	.1671
1	Canonical	11	.2273	.1354	.1591
	Non-Canonical	10	.4833	.3333	.1750
	None	12	.0556	.0278	.1346
2	Canonical	13	.1026	.1282	.0863
	Non-Canonical	16	.0313	.1563	.1094

Effect of worksheet practice

No overall differences in improvement are seen across the three conditions. There was no difference found in the amount of improvement of those children assigned to the No Practice condition and those children assigned to the Canonical Practice condition. Students in the Canonical Practice condition appear to improve slightly less than students in the No Practice condition on Clarified problems though this is not significant.

Although all first-grade students improve on Typical problems, the amount of improvement students in different worksheet conditions gain is not the same. First-grade students who received Non-Canonical Practice demonstrated significantly more improvement in their ability to successfully solve Typically worded change problems than students in the Canonical Practice condition (p=.022) or students in the No Practice condition (p=.009). These students improved much more on Typical change problems than they did on Compare problems (p=.041). In fact, they improved more on Typical problems than second-grade students. Their improvement on Clarified change problems fell between that of Typical and Compare problems and was not significantly different from either. Students in the Canonical Practice and No Practice conditions demonstrated no such difference in improvement on Typical, Clarified and Compare problems.

Second-grade students assigned to the Non-Canonical Practice condition show a different pattern of improvement than do first-grade students. They improve significantly more on Clarified problems than they do on Typical problems (p=.029). Their improvement on Compare problems is not significantly different than their improvement on either Clarified or Typical problems. As with first-grade students, children in the Canonical and No Practice conditions did not differ significantly in the amount of improvement gained on Typical, Clarified or Compare problems.

The sample of male and female children are too small to look at gender differences in the different worksheet conditions. Looking at performance on individual problems there are some small differences but nothing that suggests a pattern. The Ns are too small to be meaningful and the standard deviations are too large if one tries to examine gender differences within grades.

Although second-grade children showed improvement from pretest to posttest, the effect of the three different worksheet conditions was not significant. This is in contrast to first-grade students who clearly show differential gain scores depending on the worksheet condition to which they were assigned. Why is it that first-grade students show such large gain scores on Typically worded Change problems and second-grade students do not? It's possible that by mid-way through second grade, basic addition and subtraction facts are so well known that second-grade students treated the worksheets like rote practice. First-grade students, on the other hand, initially had a great deal of difficulty solving the non-canonical arithmetic problems.

Informal observation indicated that many of the first-grade children began using a strategy that consisted of performing the required operation on the two numbers present and filling in the calculated response on the blank line, regardless of where it appeared in the problem. For example, if a student was asked to solve 4 + __ = 9, she would add the 4 and 9 together to get 13 and write that on the blank space so the solved problem read

4 + $\underline{13}$ = 9.

CHAPTER FOUR:
CONCLUSIONS/GENERAL DISCUSSION

This work was concerned broadly with two things. Firstly, are there relatively simple changes that can be made to word problems themselves to make them easier for children to understand and to solve and what sorts of changes actually help? Secondly, are there other things that can be done to the curriculum, such as additional or different types of arithmetic practice, which will transfer to solving arithmetic word problems?

On clarifying word problems

Students answer solve-for-result (Change 1 and 2) problems with greater success than solve-for-change-set (Change 3 and 4) problems whether the wording of the problems has been clarified or is more typical of what is seen in textbooks and what has been tested by researchers. Although typically worded solve-for-change-set problems are solved correctly significantly more often than solve-for-start-set (Change 5 and 6) problems, this is not the case with Clarified problems. With Clarified problems there is no significant difference found between solve-for-change-set and solve-for-start-set problems.

Clarification of semantic and temporal relationships can improve student performance on Change problems in some cases. Students improved significantly on solve-for-start-set problems, which are the problems on which they did the most poorly in the typically worded set of problems as well as the types of Change problems that students have been shown to find the most difficult to solve.

Typically worded Change problems show the expected pattern of relative difficulty: solve-for-result problems are easier to solve than solve-for-change-set problems which are in turn easier to answer than solve-for-result problems. Students answer solve-for-result (Change 1 and 2) problems with greater success than solve-for-change-set (Change 3 and 4) problems whether the wording of the problems is typical or has been clarified.

The exception is that subtraction problems were generally solved more successfully than addition problems. In the set of problems selected, Change 1 problems were found to be easier to solve than Change 2 problems, but in both the case of solve-for-change-set and solve-for-start-set problems, the subtraction problem was solved correctly with greater frequency than the addition problem. The finding that subtraction problems are easier to solve than comparable addition problems was not expected although it could be an artifact of the problems chosen. Since there was only a single example of each type of problem, this cannot be determined without additional studies. This could also be why there were no significant difference found among the compare problems.

The attempted clarifications may have been too minor to have an effect. There were very few instances of pronoun usage and the typical problems selected may have been unintentionally clarified by conventions such as the choice of using actors of different genders. It would be interesting to do a more systematic variation of problem structure to see if problems can be made more difficult as well as less difficult. One wonders if problems can also be made more difficult by altering their wording. For example, one would expect that changing the order of information in a Change 1 problem to “Joe gave Stephanie 4 books. Before that Stephanie had 7 books. How many books does

Stephanie have now?" should increase the difficulty of the problem for young children.

On the effects of differential arithmetic practice

Additional arithmetic practice can affect improved solution of some types of word problems. Specifically, Non-Canonical arithmetic practice does seem to transfer to improvement on typically worded Change problem solution, at least among first-grade students. Practice of both types appears to help second grade students some, but not nearly so much as it helps first-grade students. There are some other intriguing trends that proved not to be significant. Neither Canonical nor Non-Canonical practice appears to differentially affect Compare problem solution compared to no practice, but this is as expected.

Why was there improvement on typically worded problems and not problems in which the wording had been clarified? First of all, it is important to note that this is not merely an effect of having more practice. Students in the Canonical Practice condition received just as many additional practice problems as students in the Non-Canonical Practice condition but show very similar improvement on Typical change problems as students in the No Practice condition. Secondly, students in the Non-Canonical Practice condition do improve significantly on Clarified change problems, it is just not significantly more improvement than students in the Canonical or No Practice conditions demonstrate on those problems.

It is actually intriguing that the Canonical practice appears to have no effect in first-grade. Students receiving no practice improved just as much as those who received Canonical practice. Second-grade students who received

Canonical practice improved on both types of Change problems, but this was not significantly more than the improvement made by students receiving no practice, even though those students' gains were not significant.

There are a number of reasons why practice may have helped first-grade students more than it helped second grade students. One possibility is that the arithmetic practice was too simple for second grade students. Since students are expected to know their addition and subtraction facts in preparation for being introduced to multi-digit arithmetic, it is possible that practice with simple addition and subtraction facts may not have been sufficiently challenging and therefore did not cause them to think about the structure of the problem. Second-grade students have had much more practice solving simple arithmetic facts than first-grade students and filling in the correct single digit answer could quickly become a matter of rote allowing students to ignore the Non-Canonical structure of the problem.

Another possibility is that second grade students may have found the arithmetic of the word problem similarly not difficult and that their scores are sufficiently high that making significant improvement on those problems is difficult. Both of these things might be addressed by creating problems, both arithmetic practice and word problems, which are more arithmetically challenging for second grade students. Initially, these studies were piloted with two-digit arithmetic problems for the second grade students, but the arithmetic was sufficiently difficult (time consuming) that the decision was made to use the same problems as were used for first-grade students. It is possible to construct 2-digit addition and subtraction problems which do not require regrouping which should be less difficult for second-grade students than the 2-digit problems which were originally piloted.

Since it is the exposure to atypical problem structures which were conjectured to assist students with word problem solution, if students were not paying attention to the structure of the practice, the projected gains would not occur. This could be addressed by making the arithmetic practice more challenging for second grade students. Perhaps if the problems did not involve regrouping or carrying, they would still provide enough challenge to require students to pay attention to the structure of the problem without becoming so time consuming as to be onerous.

Another difficulty that this research runs into is that once students were divided among three different worksheet conditions, the Ns in each condition are quite small, thus it is difficult to get significant effects. There are, however, some intriguing trends which might be explored further in a subsequent study.

One could speculate that sufficient regular practice with left-handed problems may help to mitigate the difficulties students are reported to have understanding what an equals sign represents. Students are prone to treat an equals sign as an indication to give an answer. Teachers of middle school science report that students treat an equals sign on a calculator in a similar fashion, expressing an expectation that the solution to the problem will appear when they push the equals button as opposed to using the calculator as a tool. Perhaps part of the reason that children have difficulty with the concept of equality is that in practice, given the preponderance of canonical arithmetic practice that students receive, an equals sign does represent an indication to give an answer. In that case, earlier introduction to alternative forms of arithmetic practice might help prevent that misconception from forming. Extensive practice with non-canonical arithmetic forms may also help prevent canonical arithmetic from being over-learned and may have (positive) consequences when algebra is introduced. Non-canonical arithmetic practice is in fact, a simple form of algebra that does not use a letter to represent the missing set.

Regardless of the effect or lack thereof that the two worksheet practice conditions had on word problem solution, first-grade teachers reported anecdotally that students who were in either of two worksheet conditions performed better on other classroom arithmetic tasks than their classmates who did not receive additional worksheet practice. In first-grade, in addition to learning basic addition and subtraction facts, students are taught to tell time, about calendar math and to deal with money. Review or drill of basic arithmetic facts seemed to be lacking while these special topics were being taught so perhaps that is why students in the worksheet conditions better retained basic computations skills.

Other data analyses

Identify patterns of errors

It may also be possible to analyze student numerical responses for patterns of errors that may be suggestive of the strategies used to answer each problem. Although students sometimes just guess, often they repeat one of the numbers presented in the problem. They may also perseverate at a familiar task. First-grade students in particular might perseverate at addition since it is more familiar and perhaps simpler task. This maybe especially prevalent during the pretest tasks as they had not yet been formally introduced to subtraction at that point in the school year.

Since students were asked for each question to write down the math problem that they used to get their answers, these student-reported math problems may provide further insight about the arithmetic strategies used by students to answer the word problems. Used in conjunction with numerical responses, analyzing these student-reported math problems would also allow us to identify problems in which an incorrect numerical answer is merely a computation error rather than a failure to set up a problem correctly. It would also provide a way to corroborate student-reported strategies for solving the word problems.

Analyze children's explanations of what they did

There is a rich array of student responses that could be analyzed using qualitative techniques. Since all sessions were individually audio-taped, student answers to experimenter probes could be transcribed, coded and analyzed for common threads of reasoning. Student explanations of how they solved problems may also help shed light on whether student errors are due to a comprehension failure and what type of errors they are making. As one example, the experimenter noticed informally that a number of students, in particular first-grade students, seemed to be failing to use "less" and "more" as relative terms. That is to say, students often seemed to add the terms "more" and "less" somewhat indiscriminately to a sentence without intending it to be a relative term. Children would say things such as "Andy has five" and "Andy has five more" meaning, in both cases, that Andy has a total of five grapes, not that he had five additional grapes than he had before. DeCorte and Verschaffel (1987b) also have evidence that children's understanding of what a word problem says is not always the same as what the adult who wrote the problem intended.

It has also been suggested that failures in comprehension may be related to working memory. That longer, more complex sentences may tax a young child's ability to follow what is being said simply because they are not able to hold all of it in their heads. This may also be the case with longer, more complex problems. Although children must eventually learn to understand the sometimes terse and sometimes complex language in which mathematics problems are written, it behooves us to teach them what that language is. One way of accomplishing this would be to introduce students to relatively simple problems in which the relationships are more clearly demarcated. and to teach them to recognize/understand increasingly complicated and terse problems.

Future Directions

There is evidence suggesting that it is indeed possible to affect student performance on certain types of (Change) arithmetic word problems simply by giving them additional practice on alternate forms of arithmetic problems. I suggest that this was successful because sufficient practice with alternate, non-canonical, arithmetic problems gives the students a broader variety of problems to call upon when trying to model arithmetic word problems. Pulling apart why this helps beyond speculating about it is critical to offering teachers suggestions that may actually affect their practice.

In order to answer why we need to look at several different things. How do students interpret problems? Analyzing responses and computational errors gives clues about what strategies children may be using but without giving us an understanding of why they choose to use those strategies.

DeCorte and Verschaffel's (1987b) debriefing of students offers intriguing suggestions that students sometimes do understand problems differently than the adult who wrote them intended.

Structured practice

Simply throwing additional and alternative types of arithmetic practice at students is probably not a useful suggestion. Just as we introduce addition prior to introducing concepts of subtraction, we should think about how to structure arithmetic practice. I suggest that determining a criterion for mastery of canonical math problems should be the gateway for introducing non-canonical practice and as students master the different forms that non-canonical problems should make up an increasingly larger and larger amount of their practice. Rather than giving students mixed addition and subtraction practice, arithmetic practice could be structured in a way to take advantage of a student's increasing mastery of addition and subtraction facts. As students demonstrate competence with canonical addition facts, they should be introduced to alternative forms of arithmetic sentences, perhaps beginning with left-handed solve-for-result questions (i.e., problems which present the answer first) followed by right-and left-handed solve-for-change set problems and finally right- and left-handed solve-for-start-set problems. As students develop competence at these alternative forms of arithmetic practice, they should make up an increasing percentage of the practice assigned to students to increase the pool of resources they have available.

The order in which to introduce alternative arithmetic sentences and criterion for competence prior to introducing additional or alternate forms should be tested experimentally. With first-grade students, one might want to begin with addition problems, substituting subtraction problems once they are introduced in the curriculum. At some point it seems sensible to require students to solve mixed addition and subtraction sets, thus requiring them to pay attention to which arithmetic operation is required.

Interspersing word problems among arithmetic practice may be a more effective way of teaching students to solve word problems. Wildmon, Skinner, McCurdy and Sims (1999) report that secondary school students will voluntarily choose a homework assignment with more total problems if there are simple arithmetic computation problems interspersed between the word problems over a homework assignment that contains only word problems. Students also ranked the mixed assignment as less difficult than the assignment containing onlyword problems even though the students' rate of accurate completion of the word problems on the two assignments were comparable. If one were to design mixed computation and word problem sets such that the computation was relevant to the structure of the word problems, students might be able to infer how to use the arithmetic practice to assist them with the word problems without actively being taught to do so. For example, a problem set of canonical arithmetic problems could be interspersed with Change 1 and 2 (solve-for-result) word problems or a set of arithmetic problems asking about the amount changing could be interspersed with Change 3 and 4 (solve-for-change-set) problems. Clearly, the computational practice would need to be appropriate for the students' level of understanding and word problems would need to be carefully selected such that their semantic structure matched that of the computational practice. This might assist teachers with introducing word problems of an appropriate level of complexity and requiring appropriate computational skills to students.

Practicing word problems

In addition to simplify problems so they are easier for children to understand we must also teach students to understand the 'language of mathematics'. Ultimately students must learn to parse the relatively terse and sometimes cryptic language in which mathematics problems are usually written.

There is research suggesting that both conceptually and procedurally based curricula have a positive effect on student performance, but also that both fail in certain ways (O'Rode, 2004). Children will learn what adults teach them. Children who are taught to reason conceptually about things may still have considerable difficulty with the actual computation, that is with proceduralizing what they understand. In contrast, students with a broad array of procedural tools in at their disposal may set up and solve formulas but give impractical answers like the two commonly given for the now oft-quoted school buses problem from the NAEP. One-third of a school bus is a nonsensical answer and rounding down to 31 leaves some of the passengers without transportation.

I have not at all addressed whether giving students practice on word problems themselves would assist students in solving them. It is fairly obvious to predict that it would but a better question is how might one structure such practice for optimal effect. Students are taught to look for certain key words in a problem for clues about what sort of arithmetic operation is required to solve the problem. They are taught that words such as "more" and "altogether" mean they are supposed to add and words such as "less" mean they are supposed to subtract. Unfortunately if a problem is asking the student to solve for something other than the result, the words indicating the action may not cue the arithmetic operation needed to actually solve the problem. Asking students more systematically to explain why or how they made decisions about what to do or to explain what the story problem was asking and perhaps even to write their own story problems about particular arithmetic problems might help them make the links between them.

One question that remains unanswered is why are math word problems so difficult. Given that students have better success when problems are 'simplified' is it something so simple as limitations on working memory? More complex sentence structure requires more processing resources from students. It is likely that the answer to this is actually quite complicated and that several factors each play a role in this difficulty.

APPENDIX A:

TYPICAL CHANGE PROBLEMS, ORDER A

Lauren had 8 erasers.

She lost some of them.

Now she has 6 erasers.

How many erasers did Lauren lose?

David had 11 cookies.

He gave 4 cookies to Sharon.

How many cookies does David have now?

Erica had 2 oranges.

Scott gave Erica 3 more oranges.

How many oranges does Erica have now?

Keith has 2 pencils.

How many more pencils does he need

so he has 11 pencils altogether.

Dan gave Kathy 3 acorns.

Now Dan has 7 acorns.

How many acorns did Dan have to start with?

Nick gave Sue 4 marbles.

Now Sue has 7 marbles.

How many marbles did Sue have in the beginning?

APPENDIX B:

CLARIFIED CHANGE PROBLEMS, ORDER A

Roger had 7 crayons.

Then Lori gave Roger some more crayons.

Now Roger has 9 crayons.

How many crayons did Lori give to Roger?

Jack had 11 pens.

The Jack gave Becky some pens.

Now Jack has 8 pens left.

How many pens did Jack give to Becky?

Mike had 8 apples.

Then Joyce gave Mike 4 more apples.

How many apples does Mike have now?

Justin had some bottlecaps.

Then Sherri gave Justin 5 more bottlecaps.

Now Justin has 9 bottlecaps.

How many bottlecaps did Justin have to start with?

Nancy had 6 brownies.

The Nancy gave Oliver 4 brownies.

How many brownies does Nancy have now?

Abby had some superballs.

Then Abby gave Brian 5 superballs.

Now Abby has 3 superballs left.

How many superballs did Abby have in the beginning?

APPENDIX C:

COMPARE PROBLEMS, ORDER A

Carmen caught 2 fireflies.

Jim caught 5 more fireflies than Carmen caught.

How many fireflies did Jim catch?

Joe missed 6 problems on the math test.

Joe missed 4 more problems than Marie missed.

How many problems on the math test did Marie miss?

Steven and Elizabeth went to the pet store to buy some goldfish.

Steven bought 12 goldfish.

Elizabeth bought 3 less goldfish than Steven bought.

How many goldfish did Elizabeth buy?

Kate found 2 marbles.

Kate found 8 less marbles than Billy found.

How many marbles did Billy find?

Sarah read 9 books last summer.

Sarah read 6 more books than Tim read.

How many books did Tim read last summer?

APPENDIX D:

TYPICAL CHANGE PROBLEMS, POST-TEST, ORDER A

Andy had 2 grapes.

Carol gave Andy 3 more grapes.

How many grapes does Andy have now?

Michael had 8 buttons.

He lost some of them.

Now he has 6 buttons.

How many buttons did Michael lose?

Rachel had 11 crackers.

She gave 4 crackers to Leon.

How many crackers does Rachel have now?

Martha gave Neil 3 stickers.

Now Martha has 7 stickers.

How many stickers did Martha have to start with?

Betsy gave Rob 4 seashells.

Now Rob has 7 seashells.

How many seashells did Rob have in the beginning?

Cheryl has 2 paperclips.

How many more paperclips does she need

so she has 11 paperclips altogether?

APPENDIX E:

SAMPLE CANONICAL ARITHMETIC WORKSHEET

Name ______________________________ Date ______________

Solve the problem by filling in the blank.

17 - 9 = __	9 + 7 = __
4 + 3 = __	1 + 3 = __
5 + 1 = __	9 + 8 = __
3 - 2 = __	10 - 7 = __
15 - 6 =	12 - 9 = __
14 + 3 = __	3 + 7 = __
14 - 9 = __	7 + 3 = __
7 + 4 =	

APPENDIX F:

SAMPLE NONCANONICAL ARITHMETIC WORKSHEET

Name ________________________________ Date ________________

Solve the problem by filling in the blank.

__ - 9 = 8

__ + 3 = 7

6 = __ + 1

1 = 3 - __

9 = __ - 6

__ + 3 = 11

__ - 9 = 5

11 = 7 + __

9 + __ = 16

1 + __ = 4

17 = __ + 8

3 = 10 - __

__ - 9 = 3

10 = __ + 7

7 + __ = 10

REFERENCES

Blanco, S. (1997). personal correpondence.

Briars, D. J., & Larkin, J. H. (1984). An integrated model of skill in solving elementary word problems. *Cognition and Instruction, 1,* 245-296.

Cardelle-Elawar, M. (1992). Effects of teaching metacognitive skills to students with low mathematical ability. *Teaching and Teacher Education: An International Journal, 8,* 3-14.

Carpenter, T. P. (1985). Learning to add and subtract: An exercise in problem solving. In E. A. Silver (Ed.), *Teaching and learning mathematical problem solving: Multiple research perspectives (pp. 17-40).* Hillsdale, NJ: Lawrence Erlbaum.

Carpenter, T. P., & Moser, J. M. (1982). The development of addition and subtraction problem-solving skills. In T. P. Carpenter, J. M. Moser, & T. A. Rombers (Eds.), *Addition and subtraction: A cognitive perspective.* Hillsdale, NJ: Lawrence Erlbaum.

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

Carpenter, T. P., Ansell, E., Franke, M. L., Fennema, E., & Weisbeck, L. (1993). Models of problem solving: A study of kindergarten children's problem-solving processes, *Journal for Research in Mathematics Education, 24(5),* 428-441.

Carpenter, T. P., Fennema, E., Peterson, P. L., Chiang, C. P., & Loef, M. (1989). Using knowledge of children's mathematics thinking in classroom teaching: An experimental study. *American Educational Research Journal, 26,* 488-531.

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, January,* 27 - 39.

Carpenter, T. P., Hiebert J. & Moser, J. M. (1983). The effect of instruction on children's solutions of addition and subtraction word problems. *Educational Studies in Mathematics, 14,* 56-72.

Carpenter, T. P., Moser, J. M., & Bebout, H. C. (1988). Representation of addition and subtraction word problems. *Journal for Research in Mathematics Education*, *19(4),* 345-357.

Chi, M. T. H., Glaser, R., & Farr, M. J. (Eds.) (1988). *The nature of expertise.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Cummins, D. D. (1991). Children's interpretations of arithmetic word problems. *Cognition and Instruction, 8,* 261-289.

Cummins, D. D., Kintsch, W., Reisser, K., & Weimer, R. (1988). The role of understanding in solving word problems. *Cognitive Psychology, 20,* 439-462.

DeCorte, E., & Verschaffel, L. (1981). Children's solution processes in elementary arithmetic problems: Analysis and improvement. *Journal of Educational Psychology*, *73 (6),* 765-779.

DeCorte, E. & Verschaffel, L. (1985). Beginning first graders' initial representation of arithmetic word problems. *Journal of Mathematical Behavior, 4,* 3-21.

DeCorte, E., & Verschaffel, L. (1987a). The effect of semantic structure on first graders' strategies for solving addition and subtraction word problems. *Journal for Research in Mathematics Education, 18(5),* 363-381.

DeCorte, E., & Verschaffel, L. (1987b). Using retelling data to study young children's word-problem solving. In D. Rogers & J. Sloboda (Eds.), *Cognitive processes in mathematics (pp. 42-59).* Oxford, England: Clarendon Press.

DeCorte, E. & Verschaffel, L. (1991). Some factors influencing the solution of addition and subtraction word problems. In K. Durkin & B. Shire (Eds.), *Language and mathematical education (pp. 117-130).* Milton Keynes: Open University Press. (91/6).

DeCorte, E., Verschaffel, L., & DeWin, L. (1985). Influence of rewording verbal problem on children's problem representations and solutions. *Journal of Educational Psychology, 77(4)*, 460-470.

DeCorte, E., Verschaffel, L., & Pauwels, A. (1990). Influence of the semantic structure of word problems on second graders' eye movements. *Journal of Educational Psychology, 82(2),* 359-365.

Dossey, J. A., Mullis, I. V. S., Lindquist, M. M., & Chambers, D. L. (1988). *The mathematics report card: Are we measuring up?* Princeton, NJ: Educational Testing Service.

Fuson, K. C. (1990a). Conceptual structures for multiunit numbers: Implications for learning and teaching multidigit addition, subtraction, and place value. *Cognition and Instruction, 7(4),* 242-403.

Fuson, K. C. (1990b). Issues in place-value and multidigit addition and subtraction learning and teaching. *Journal for Research in Mathematics Education, 21(4),* 273-280.

Fuson, K. C. (1992a). Research on whole number addition and subtraction. In D. Grouws (Ed.), *Handbook of research on mathematics teaching and learning (pp. 243-275).* New York: Macmillan.

Fuson, K. C. (1992b). Research on learning and teaching addition and subtraction of whole numbers. In G. Leinhardt, R. T. Putnam, & R. Hattrup (Eds.), *Analysis of arithmetic for mathematics teaching* . Hillsdale, NJ: Lawrence Erlbaum.

Fuson, K. C., & Kwon, Y. (1991). Chinese-based regualr and European-based irregular systems of number words: The disadvantages for English speaking children. In K. Durkin & B. Shire (Eds.), *Language in mathematics education (pp. 211-226).* Milton Keynes, G. B. : Open University Press.

Fuson, K. C., Stigler, J. W., & Bartsch, K. (1986). Grade placement of addition and subtraction topics in China, Japan, the Soviet Union, Taiwan, and the United States. *Journal for Research in Mathematics Education, 19,* 449-458.

Gelman, R., & Gallistel, C. R. (1978). *The child's understanding of number*. Cambridge, MA: Harvard University Press.

Hatano, G. (1982). Learning to add and subtract: A Japanese perspective. In TP Carpenter, Moser & Romberg (Eds.) *Addition and subtraction: A cognitive perspective (pp. 211-223).* Hillsdale, NJ: Erlbaum.

Hegarty, M., Mayer, R. E., & Monk, C. A. (1995). Comprehension of arithmetic word problems. *Journal of Educational Psychology, 85,* 18-32.

Hegarty, M., Mayer, R. E., & Green, C. (1992). Comprehension of arithmetic word problems: Evidence from students' eye fixations. *Journal of Educational Psychology, 84,* 76-84.

Hess, R. D., & Azuma, H. (1991, December). Cultural support for schooling: Contrasts between Japan and the United States. *Educational Researcher, 20(9),* 2-8.

Hess, R. D., Chang & McDevitt, T. M. (1987) Cultural variations in family beliefs about children's performance in mathematics: Comparisons among Peoples Republic of China, Chinese-American, and Caucasian-American Families, *Journal of Educational Psychology, 79,* 179-188.

Hudson, T. (1983). Correspondence and numerical differences between disjoint sets. *Child Development, 54*, 84-90.

Kintsch, W., & Greeno, J. G. (1985). Understanding and solving word problems. *Psychological Review, 92,* 109-129.

Kouba, V. L., Brown, C. A., Carpenter, T. P., Lindquist, M. M., Silver, E. A., & Swafford, J. O. (1988). Results of the fourth NAEP assessment of mathematics: Number, operations and word problems. *Arithmetic Teacher, 35(8),* 14-19.

LaPointe, A. E., Meade, N. A., & Phillips, G. W. (1989). *A world of differences: In international assessment of mathematics and science* . Princeton, NJ: Educational Testing Services.

Lewis, A. B. (1989). Training students to represent arithmetic word problems. *Journal of Educational Psychology, 79,* 363-371.

Lewis, A. B., & Nathan, M. J. (1991) A framework for improving students' comprehension of word arithmetic and word algebra problems. In L. Birnbaum (Ed.), *Proceedings of the International Conference on the Learning Sciences (pp. 305-314).* Charlottesville, VA: Association for the Advancement of Computing in Education.

Lewis, A. B., & Mayer, R. E. (1987). Students' miscomprehension of relational statements in arithmetic word problems. *Journal of Educational Psychology, 79*, 363-371.

Mautone, J. A. (1999). *Children's processing of arithmetic word problems: Evidence from Eye movements.* Unpublished senior honors thesis, Cornell University.

Mayer, R.E. (1981). Frequency norms and structural analysis of algebra story problems into families, categories and templates. *Instructional Science, 10,* 135-175.

Mayer, R. E. (1982). Memory for algebra story problems. *Journal of Educational Psychology, 74,* 199-216.

Mayer, R. E. (1985). Mathematical Ability. In R. J. Sternberg (Ed.), *Human Abilities: An Information Processing Approach (pp. 127-150).* New York, NY: Freeman.

Mayer, R. E. (1998). Cognitive, metacognitive, and motivational aspects of problem solving. *Instructional Science, 26,* 49-63.

Mayer, R. E. (1991). *Thinking, Problem Solving, Cognition (2nd ed.).* New York: Freeman.

Mayer, R. E., & Hegarty, M. (1996). The process of understanding mathematical problems. In R. J. Sternberg & T. Ben-Zeev (Eds.), *The nature of mathematical thinking (pp. 29-53)*. Mahwah, NJ: Lawrence Erlbaum Associates.

Mayer, R. E., Larkin, J. H., & Kadane, J. B. (1984). A cognitive analysis of mathematical problem solving ability. In: Sternberg, R. (Ed.), *Advances in the Psychology of Human Intelligence,* Lawrence Erlbaum, Hillsdale, NJ

Mayer, R. E., Lewis, A. B., & Hegarty, M. (1992). Mathematical misunderstandings: Qualitative reasoning about quantitative problems. In J. I. D. Campbell (Ed.) *The nature and origins of mathematical skills (pp. 137-153).* Amsterdam: Elsevier.

Miura, I. T. (1987). Mathematics achievement as a function of language. *Journal of Educational Psychology, 79(1)*, 79-82.

Miura, I. T., Kim, C. C., Chang, C-M., & Okamoto, Y. (1988). Effects of language characteristics on children's cognitive representations of number: Cross-national comparisons. *Child Development, 59,* 1445-1450.

Miura, I. T., Okamoto, Y., Kim, C. C., Steere, M., & Fayol, M. (1993). First graders' cognitive representation of number and understanding of place value: Cross-national comparisons –– France, Japan, Korea, Sweden and the United States. *Journal of Educational Psychology, 85(1),* 24-30.

Polya, G. (1965). *Mathematical discovery on understanding, learning and teaching problem solving, Vol. 2.* London: Wiley.

O'Rode, N. (2004). *Making sense of simple equations: How students using IMP and CMP curricula understand equivalence.* Paper presented at the annual meeting of the American Educatution Research Association, San Diego.

Resnick, L. B., & Omanson, S. F. (1987). Learning to understand arithmetic. In R. Glaser (Ed.) *Advances in Instructional Psychology, Vol. 3 (pp. 109-151).* New York: Academic Press.

Reusser K. (1989). *Textual and situational factors in solving mathematical word problems.* (Research Report 7). Bern, Switzerland: University of Bern, Department of Educational Psychology.

Reusser K. (1990). From Text to Situation to Equation: Cognitive Simulation of Understanding and Solving Mathematical Word Problems. In H. Mandl, E. De Corte, N. Bennet & H.F. Friedrich (Eds.), *Learning and Instruction, European Research in an International Context, Vol. II.* New York: Pergamon Press.

Riley, M. S., & Greeno, J. G. (1988). Developmental analysis of understanding language about quantities and of solving problems. *Cognition and Instruction, 5,* 49-101.

Riley, M. S., Greeno, J. G., & Heller, J. I. (1983). Development of children's problem solving abilities in arithmetic. In H. P. Ginsburg (Ed.), *The development of mathematical thinking (pp. 153-196).* San Diego: Academic Press.

Robitaille, D. F. & Garden, R. A. (1989). *The IEA study of mathematics II: Contexts and outcomes of school mathematics,* Pergammon.

Robitaille, D. F. & Travers, K. (1992). International studies of Achievement in Mathematics. In D. Grouws (Ed.), *Handbook of research on mathematics teaching and learning.* New York: Macmillan.

Sherwood, S. (1997). personal correspondence.

Song, M. J., & Ginsburg, H. P. (1987). The development of informal and formal mathematics thinking in Korean and U. S. children. *Child Development, 58,* 1286-1296.

Stern, E. (1993). What makes certain arithmetic word problems involving comparison of sets so difficult for children. *Journal of Educational Psychology, 85,* 7-23.

Stern, E., & Lehrndorfer, A. (1992). The role of situational context in solving word problems. *Cognitive Development, 2,* 259-268.

Sternberg, R. J., & Frensch, P. A. (Eds.) (1991). *Complex Problem Solving: Principles and Mechanisms.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Stevenson, H. W., Lee, S. Y., & Stigler, J. W. (1986). Mathematics achievement of Chinese, Japanese, and American children. *Science, 231,* 693-699.

Stevenson, H. W., Lummis, M., Lee, S. Y., & Stigler, J. W. (1990). *Making the grade in mathematics: Elementary school mathematics in the United States, Taiwan, and Japan*. Reston, VA: National Council of Teacher of Mathematics.

Stevenson, H. W. & Stigler, J. W. (1992). *The Learning Gap.* New York: Summit Books.

Stigler, J. W. (1988). The use of verbal explanation in Japanese and American classrooms. *Arithmetic Teacher, 36(2),* 27-29.

Stigler, J. W., Fuson, K. C., Ham, M., & Kim, M. S. (1986). An analysis of addition and subtraction word problems in American and Soviet elementary textbooks. *Cognition and Instruction, 3,* 153-171.

Stigler, J. W., Lee, S. Y., Lucker, G. W., & Stevenson, H. W. (1982). Curriculum and achievement in mathematics: A study of elementary school children in Japan, Taiwan, and the United States. *Journal of Educational Psychology, 74,* 315-322.

Stigler, J. W., Lee, S-Y., & Stevenson, H. W. (1990). *Mathematical knowledge of Japanese, Chinese and American elementary school children.* Reston, VA: National Council of Teachers of Mathematics.

Verschaffel, L. (1984). Representatie- en oplossingsprocessen van eersteklassers bij aanvankelijke redactie-opgaven over optellen en aftrekken. Dissertatie, Katholieke Universiteit Leuven.

Verschaffel, L., DeCorte, E., & Pauwels, A. (1992). Solving compare problems: An eye movement test of Lewis and Mayer's consistency hypothesis. *Journal of Educational Psychology, 84(1),* 85-94.

Wertheimer, M. (1959). *Productive Thinking (Enlarged Ed.).* New York:Harper & Row.

Wildmon, M. E., Skinner, C. H., McCurdy, M. & Sims, S. (1999). Improving secondary students' perception of the "dreaded mathematics word problem assignment" by giving them more word problems. *Psychology in the Schools, 36(4),* 319-325.