Children's Recall of Arithmetic Word Problems

By

Denise Dellarosa, Rhonda Weimer, and Walter Kintsch

University of Colorado

This work was supported by National Science Foundation Grant No. BNS-8309075 to W. Kintsch and J.G. Greeno.

Abstract

Word problems are notoriously difficult to solve. Children typically perform 10% to 30% worse on word problems than on comparable problems presented in numeric format. Moreover, characteristic errors tend to be committed with certain problem types. Researchers have tried to explain these errors by making inferences about the types of problem representations children build during solution attempts. In the present study, we took a more direct approach. We chose easy and hard problems from each of three standard problem types and required second grade children to solve and recall them. We found that (1) children tended to spontaneously transform difficult problems into simpler ones during solution attempts, (2) these transformations often resulted in structurally incorrect problem representations, and (3) the errors they committed were directly related to the problem representation they had built. Often these "errors" were the correct answers to the problem structures they had incorrectly built.

Children's Recall of Arithmetic Word Problems

A major goal of education is to produce competent problem solvers. We educate our children not simply to produce well-stocked repositories of information, but instead to provide them with the knowledge and tools they will need to solve problems and enrich their lives. Typically, these problems will present themselves as verbally described situations, and often these situations will require computations to achieve their solution. When presented in the classroom, simple instances of these problem situations are called arithmetic word problems.

Word problems are notoriously difficult to solve. Children typically perform 10% to 30% worse on word problems than on comparable problems presented in numeric format (Carpenter, Corbitt, Kepner, Lindquist, & Reys, 1980). Results such as these suggest that the difficulty children encounter with such problems does not stem from an inability to manipulate numbers successfully. The true sources of such difficulty, however, is not well understood.

The search for the source or sources of word problem difficulty has generally taken two paths (Briars & Larkin, 1984). The first involves analyses of problem features that contribute to solution difficulty. These features include the number of words in the problem, the presence or absence of words that cue arithmetic operations, the size of the numbers involved, whether or not actions are involved and the type of equations embedded in the problems. While these factors do influence problem difficulty, they do not account for as much variation in solution success as one might expect, and their influence diminishes with age (Carpenter, Hiebert, & Moser, 1981; Carpenter & Moser, 1982).

The second approach explains differences in problem difficulty by analyzing the cognitive processes required to solve them (Briars & Larkin, 1984; Kintsch & Greeno, 1985; Riley, Greeno, & Heller, 1983). This is an interactive approach;

it assumes that problem features interact with the cognitive capacities of the child to produce characteristic levels of problem difficulty. A feature proves troublesome only insofar as the capacity required to process that feature is not yet possessed by the child.

The interactive approach to problem difficulty begins with a task analysis, that is, an analysis of the processes and knowledge required for successful task performance. A gross level of analysis would at least include:

1. Comprehension of the problem text, including
   (a) understanding the words
   (b) understanding the semantics of each sentence
2. Comprehension of the problem situation, including
   (a) activating world knowledge concerning the situation
   (b) activating knowledge about set relations and mathematics
   (c) using information gleaned from 1
3. Choosing or deriving a solution strategy based on 2
4. Executing the chosen strategy correctly

Problems presented in numeric format require only process #4 in order to be solved, hence their relative simplicity. Once the proper addition and subtraction algorithms have been mastered (in itself no mean feat), numeric problems require no more than executing the algorithm specified by the arithmetic signs in the problem for their successful solution.

Word problems, on the other hand, require three additional processes, all of which depend heavily on the child's previously acquired knowledge. Many theorists assume that the bulk of learning, and hence problem difficulty, lies in process #2, and specifically in #2b. As children acquire knowledge about set relations and mathematics, problems become simpler to solve. Briars and Larkin, for example, argue convincingly that several standard problem types prove difficult for children to solve because they require (a) knowledge that one object can be a member of both a set and a superset simultaneously, (b) knowledge that processes can be undone or done backward in time, and (c) that the roles of subsets can be exchanged while deriving a solution. These three pieces of knowledge all describe logical relations among sets that, theoretically, very young children do not yet possess.

However plausible this account of problem difficulty, it at times has found itself at odds with empirical results. Several studies have uncovered a surprising amount of sophistication among children in their understanding of arithmetic operations and logical relations among sets. Hudson (1983), for example, presented kindergartners and first graders problems such as the following along with pictures of the described situation:

(a) There are 5 birds and 3 worms.
How many more birds are there than worms?

(b) There are 5 birds and 3 worms.
How many birds won't get a worm?

Notice that problem (b) is merely a reformulation of problem (a) in a simpler linguistic form. The arithmetic and logical set relations in both are identical. Problems of type (a) are notoriously hard for children up to the third grade. Indeed, only 39% of Hudson's first graders obtained the correct answer to problems of this type. (The most common error was to give the cardinality of the larger set as the answer because of the word 'more' in the question.) In contrast, however, 79% of the first graders got problems of type (b) correct. Moreover, the majority of children solved problems of the latter type not by simple pairing of objects from the two sets, but by counting out a subset of the larger set equal to the smaller and then counting the remaining elements. These results were surprising because they showed that young children fail comparative problems not because they misunderstand correspondence set relations or lack counting skills, but because they find questions in comparative linguistic form to be difficult to understand. The solution strategies even kindergartners used revealed an implicit understanding that a one-to-one correspondence between two sets necessarily exists if the two can be counted out to the same number. Similarly surprising results were observed by DeCorte & Vershaffel (1984), who produced substantial increases in solvability of typically unsolvable problems simply by making the semantic relations among sentences within the problems more explicit. Here again, a linguistic manipulation was found to was found to improve arithmetic performance.

These results imply that the representations children build are influenced by the linguistic structures of the problems they encounter. Children fail to understand the problem situation described by a complex linguistic structure. When this situation is not well understood, the mathematical and logical knowledge required to solve the problem is less likely to be activated, and incorrect solutions result. On the other hand, simple linguistic structures allow the problem situation to be readily understood, and hence are more likely to cue correct solution strategies. The development of arithmetic problem solving expertise and linguistic competence are therefore interdependent skills.

Problems differ not just in linguistic difficulty, however; the problem situations they describe may differ in complexity as well. For example, compare-type problems are never solved as easily as certain other problem types regardless of how simply the former are worded. This is presumably because a comparison situation is simply more difficult for a child to conceptualize than (e.g.) a combine situation, wherein two quantities are simply combined to yield an answer. This is also true when the compare problem also requires addition. Problem structures also differ in difficulty depending on what is unknown and has to be computed. For example, conceptualizing a situation wherein the cardinal of a result set is unknown is generally easier for children then conceptualizing one in which a start set cardinal is unknown.

A logical strategy to adopt when a problem appears too complex to solve is to simplify it, linguistically or otherwise. For example, Riley, et al., observed that when children were asked to recall problems they had solved incorrectly, the problems tended to be misrecalled as simpler ones that could be solved. These results suggest that when children find a problem difficult to understand, they may try to transform it's structure into simpler ones that know how to solve. If this were the case, then a relation should exist between the answer produced and the problem situation recalled later. Unfortunately, the authors did not make this comparison. It was with this aim in mind that the present study was conducted.

We required children to recall, solve, or both solve and recall problems of two levels of difficulty. The recall protocols provided direct, first-hand information concerning the children's representations of the problems in memory both without and following a solution attempt. We predicted the following: First, that more recall errors would be observed for difficult problems than easy problems; second, that the nature of these errors would be in the form of simplifying the problems to an easier type of problem; and third, that errors in recall would be associated with solution failure and incorrect choice of solution strategies.

METHOD

Subjects. Thirty second grade children from the Boulder Valley School District served as subjects in the study. Children were recruited through notices distributed to second grade classes and through newspaper advertisements. Parental consent was obtained prior to each session. Each child was paid $5.00 for his or her participation. Each session lasted approximately 1/2 hour.

Materials. Thirty word arithmetic problems served as materials in the study. These problems were divided equally among three problem types: Change problems (CH), Combine problems (CB), and Compare problems (CP). These problem types represented three distinct linguistic structures as well as requiring distinct strategies for their solutions (Carpenter & Moser, 1982). Within each problem type were an equal number of easy problems and hard problems. The level of difficulty of each problem was determined by proportion correct performance by second grade children tested by Riley, Greeno, & Heller (1983). Examples of the problems used are presented in Appendix A, along with their respective performance levels in the Riley et al. study. Five instances of each of these six problem groups were used. They were as follows: Easy problems--Change 1, Combine 1, and Compare 4; Hard problems--Change 5, Combine 2, Compare 5.

Procedure. Experimental sessions took place in the child's home. The experimenter tape recorded all sessions.

The child's name, age, and grade level were recorded on the tape, and he or she was instructed as to the nature of the task. The child was told that the experimenter would read a problem aloud, and then would ask him or her to do one of three things: Repeat the problem aloud (RECALL ONLY), solve the problem (SOLVE ONLY), or solve the problem and then repeat it (SOLVE & RECALL). Examples were given of the three tasks. The child was then given a practice session using one problem from each of the six problem groups. He or she was required to recall two of the problems, solve two of the problems, and solve and recall two. Following the practice session, the instructions were reiterated, and the child was questioned to ensure that he or she understood the task. Following this, the experimental session was begun.

Of the remaining four problems in each problem group, one was assigned to the RECALL ONLY condition, one was assigned to the SOLVE ONLY condition, and the remaining two were assigned to the SOLVE & RECALL condition. The RECALL ONLY and SOLVE ONLY conditions were essentially control conditions providing baseline performance for the recall and solution measures. Twice as many problems were assigned to the SOLVE & RECALL condition because we were primarily interested in recall characteristics following both correct and incorrect solutions.

Presentation order was randomized for each subject. Responses were prompted by requesting the child to repeat the problem or to answer it. In addition, the experimenter provided the names of the children mentioned in the story. This was done because pilot work had indicated that children immediately forgot the names used in the problems and used up considerable memory resources in attempts to retrieve them. Care was taken not to prompt problem structure while providing names. For example, an important aspect of COMPARE 5 problems is that they contain a pronoun instead of a proper name in the second line. In these cases, when the child failed to use a pronoun and requested a name, he or she was required to indicate whether the desired name was the first or second person mentioned in the problem. Finally, repetitions of problems were given if the child forgot substantial portions of the problem due to inattention.

RESULTS

Unless otherwise noted, rejection probability for all statistical tests was .05. Statistically significant interactions were followed by simple effects tests (Keppel 1973); statistically significant main effects and simple main effects involving more than one mean were analyzed using Tukey's HSD test. All data were scored from tape recordings made during the sessions.

Solution Performance. Presented in Table 1 are the mean proportion correct solutions for each of the six problem groups for the present study and that of Riley et al. As is apparent, performance levels in the present study compare favorably to those of Riley et al., with the possible exception of COMPARE 5. Performance in the present study was much higher on this problem than in Riley et al, probably because our children were tested much later in the school year. An analysis of variance was performed on these data using as variables level of difficulty (Easy and Hard), and problem type (Change, Combine, and Compare), with repeated measures on both variables. The analysis returned a single significant result, that of the main effect of difficulty, $F(1,29) = 29.59$, $MS_e = 2.03$, $p < .0001$, indicating that more easy problems were correctly solved than hard problems, regardless of problem type.

Table 1

Mean Proportion Correct Solutions To Word Problems

Level of Difficulty	Problem Type: Change	Combine	Compare
Easy	.89 (1.00)	.93 (1.00)	.79 (.80)
Hard	.61 (.75)	.68 (.55)	.68 (.35)

Note: Proportions are based on performance on three problems for thirty subjects. The maximum possible score for each child on each problem type was three and the minimum zero. Numbers in parentheses are the mean proportion correct solution performance from Riley, Greeno, and Heller (1983).

Solution latencies. Also recorded was the times children required to solve each of the problems. The distribution of these data was highly positively skewed, so they were subjected to a normalizing logarithmic transformation prior to analyzing them. The mean time, in seconds, required to solve each of the problem types, are presented in Table 2, along with raw means, and the means conditionalized on correct solutions. The data from each subject was averaged across three instances of each problem type.

The first thing to notice is that the trends in the data are the same regardless of whether the latencies are raw latencies, log transformed latencies, or conditionalized latencies. The log transformed latencies were analyzed via an analysis of variance, using problem difficulty (Easy and Hard problems) and problem type (Change, Combine, and Compare) as within variables. The main effects of these two variables were significant, $F(1,29)=10.81$, $MS_e=.19$, $p<.01$, and $F(2,58)=7.92$, $MS_e=.24$, $p<.01$, respectively. Children required significantly more time to solve hard problems than to solve easy problems. Moreover, they took longer to solve Compare problems than to solve either Combine or Change problems; the latter two types required an equivalent amount of time. The longer latencies to hard problems do not seem to be due to errors. Although the mean solution times decrease when error trials are removed from them, indicating that errors are associated with longer solution times, the relative rankings of the mean remain the same: Easy problems still require less time than hard problems, and Compare problems still require more time to solve than either of the other two types. (Conditionalized latencies could not be analyzed because many children made errors on all three instances of the harder problem types resulting in several empty cells.)

Table 2

Solution Latencies: Mean Number of Seconds to Produce an Answer

	Problem Type		
	Change	Combine	Compare
Easy Problems			
Raw	4.21	3.69	5.86
Conditionalized	3.44	3.68	5.24
Log Transformed	3.32	3.27	3.73
Hard Problems			
Raw	5.47	5.80	6.40
Conditionalized	4.76	4.79	6.10
Log Transformed	3.53	3.66	3.79

Note: Means in each cell are based on an average of latencies for three problems from each of thirty children. Conditionalized latencies have error trials removed from them.

Word Recall. As a baseline measure of subject's abilities to perform the recall task, the mean proportion of words in each problem recalled correctly was computed and analyzed. The means for each of the problem groups are presented in Table 3. These means represent merely the subject's abilities to recall words that were used in the problem, and not accurate recall of problem structure. For example, a COMBINE 2 problem could have been recalled as a COMBINE 1 problem and still produce a perfect recall score since all of the correct words would have been repeated. Structure recall was analyzed separately. Note also that the relative rankings of the means in the solve and recall condition remain constant even after conditionalizing recall on correct solutions. Therefore, the differences to be discussed below are not due to disproportionate solution errors on hard problems.

Mean word recall was subjected to an analysis of variance using the following variables: level of difficulty (Easy and Hard), problem type (Change, Combine, and Compare), and recall condition (Recall Only and After Solution), with repeated measures on all three variables. Several effects were significant. First, the problem types were found to differ in recallability, $F(2,58) = 59.70$, $MSe = .014$, $p < .0001$. Tukey's test indicated that more words tended to be recalled from the Combine problems than from the Compare problems, which in turn tended to be better recalled than the Change problems. The required difference was .038, and the obtained differences were: CH vs CB = .169, CH vs CP = .090, and CB vs CP = .079. Poor recall of Change problems was almost certainly due to two factors. First, unlike the other problems, Change 5 has four lines instead of three. (There are six types of Change problems, four of which have four lines, and only two of these four prove difficult for children to solve.) Children tended to fail to recall large portions of either the third or fourth line of these problems, suggesting that their short term retention is strained beyond three lines. The second factor most likely contributing to poor recall of Change problems is the fact that these problems contain references to time, such as 'then' and 'now'. While these time-words may prove helpful in determining set roles such as the start set and result set, they tended to be forgotten during recall. This may be because the time course of events in these problems is implicitly understood, making time-words redundant and easily forgotten.

The main effect of difficulty was significant, $F(1,29) = 18.24$, $MSe = .03$, $p < .001$. However, this effect was modified by a significant interaction with recall condition, $F(1,29) = 10.31$, $MSe = .01$, $p < .01$. Simple effects tests indicated that subjects recalled less of the hard problems following a solution attempt than they did of the easy problems, $F(1,58) = 9.37$, $MSe = .019$, $p < .01$. No other simple effect was significant, although a tendency to recall fewer words from the hard problems following a solution attempt than was produced in the recall only condition was observed, $F(1,58) = 3.61$, $MSe = .011$, $.05 < p <<$

.10. This pattern of results suggests that subjects used so few cognitive resources in solving easy problems that recalling them following a solution attempt was as effortless as recalling them alone. Hard problems, on the other hand, required substantial cognitive resources to solve, and hence retention of the problem text suffered. The most important aspect of these findings, however, is that children were found to be capable of performing a verbatim recall task, and that the resources required for this are not overwhelming.

Table 3

Mean Proportion of Words Recalled From Word Problems

	Problem Type		
Easy Problems	Change	Combine	Compare
Recall Only	.73	.88	.80
After Solution	.76 (.77)	.89 (.89)	.80 (.84)
Hard Problems			
Recall Only	.65	.85	.78
After Solution	.59 (.66)	.81 (.85)	.72(.74)

Note: Proportions in the recall only condition are based on one observation from each of thirty subjects. Proportions after solutions are based on two observations from each of thirty subjects. Each observation was computed by counting the number of words recalled by a subject from a particular problem and dividing it by the total possible number of words that could be recalled from that problem. Numbers in parentheses represent recall proportions conditionalized on correct solutions.

Recall latencies. To obtain some idea of how long it took children on the average to recall the problems used, the number of seconds required to recall each problem was recorded. These recall times were then divided down by the number of words recalled for each problem since the problems differed in the number of words they contained. Finally, these data were trimmed to remove recall times in excess of three standard deviations beyond the mean; in the solve and recall condition they were then averaged across the two occasions within each problem group. The mean recall rate for each problem group (i.e., seconds per word) in both the recall only and solve and recall conditions are presented in Table 4. An analysis of variance was conducted on these data using as variables recall condition (Recall Only and After Solution), problem difficulty (Easy and Hard problems), and problem type (Change, Combine, and Compare). The main effects of problem difficulty and problem type, and their interaction were significant, $\underline{F}(1,29)=18.69$, $MS\underline{e}=.21$, $p < .001$, $\underline{F}(2,58)=16.64$, $MS\underline{e}=.24$, $p < .001$, $\underline{F}(2,58)=9.43$, $MS\underline{e}=.11$, $p < .001$, respectively. The problem types were found to differ significantly from each other at both levels of difficulty, $\underline{F}s(2,58)=6.41$ and 3.26, respectively, $MS\underline{e}=.18$, $p < .01$ and .05, respectively. Tukey's test indicated that COMBINE 1 took the least amount of time to recall, followed by CHANGE 1 and finally COMPARE 4. The ordering of the difficult problems was different, however, Here, COMPARE 5 was found to take the most time, while the other two did not differ. The change in ordering appeared to be caused primarily by the COMBINE problems. COMBINE 2 was found to require significantly more time to recall than COMBINE 1, $\underline{F}(1,87)=18.52$, $MS\underline{e}=.15$, $p < .001$, while the time it took to recall the CHANGE and COMPARE problems did not vary with their level of difficulty, $\underline{F}s < 1$.

It is interesting to note that while the time rate required to recall the problems did not differ in the recall only and after solution conditions, they total number of words recalled was fewer following a solution. A straightforward interpretation of these results is that a constant time is required to recall words from a particular problem type regardless of the number of words still available in memory. It is the difficulty of the particular structure that determines how long it will take to retrieve a part of the structure, and not the number of its elements per se.

Table 4

Recall latencies: Average number of seconds per word recalled

	Problem type		
Easy Problems	Change	Combine	Compare
Recall Only	1.06	0.92	1.43
After Solution	1.12	0.83	1.28
Hard Problems			
Recall Only	1.11	1.26	1.47
After Solution	1.21	1.34	1.51

Note: Each mean in the recall only condition is based on one observation from each of thirty children. Each mean in the after solution condition is based on an average of two occurrences from each of thirty children. Outliers have been removed.

Structural errors in recall. We have argued that a major strategy that children employ when faced with linguistically or mathematically complex problems is that of simplification. In particular, we predicted that more recall errors should be observed for difficult problems than easy problems, and that the nature of these errors should be in the form of simplifying their structures. This implies that when difficult problem structures are misrecalled, they should be misrecalled as simpler problem structures. To test these predictions, the number and types of errors made while recalling problem structure were scored, and are presented in Table 5. If such a simplification bias existed, there should be more observations in quadrant III of this table than in any of the other quadrants. This is clearly the case. Quadrants I and IV are relatively empty, indicating that subjects rarely misrecall easy problems as hard problems, nor do they misrecall hard problems as other types of hard problems. Two tendencies are apparent in quadrants II and III. First, when errors were made, children exhibited a slight tendency to misrecall all problems as COMBINE 1 problems, suggesting that COMBINE 1 is the simplest type of problem structure for them. Second, and more importantly, when children misrecalled hard problems they tended overwhelmingly to misrecall them as simpler problems, and moreover as simpler problems of the same type. Thus, CHANGE 5 tended to be misrecalled as CHANGE 1, COMBINE 2 as COMBINE 1, and COMPARE 5 as COMPARE 4. These differences were statistically significant, $X^2(30)=145.14$, $p < .001$.

Perhaps the most interesting error type is the overwhelming tendency to misrecall COMPARE 5 as COMPARE 4. Note that in order to commit this error, it is necessary to perform several transformations to the problem text. Specifically, it is necessary to switch the position of the names in the second line, and to change the word 'more' to the word 'less'. However, at an arithmetic level, these two problems are identical, their equations both being:

$$a - b = ?$$

It is particularly relevant to our discussion here to note that children, when making errors in recalling this particular problem type, tended to maintain the correct arithmetic structure while simplifying the verbal structure. Note that the word 'less' in COMPARE 4 is consistent with the operation to be performed (i.e., subtraction), while the word 'more' in COMPARE 5 is inconsistent with that operation.

The same cannot be said, however, for either CHANGE 5 or COMBINE 2. The most frequent recall errors for these types were the following:

Change 5 Jane had 5 marbles. (Transfer set in original problem)
Karen gave her 8 more marbles. (Result set in original problem)
How many marbles does Jane have now?

Combine 2 George has 9 marbles. (SuperSet in original problem)
Mary has 4 marbles. (Known Subset in original problem)
How many marbles do they have altogether?

In both of these cases, the verbal structure was simplified, but at the risk of changing the arithmetic structure of these problems. An important question is whether these recall errors were in any way related to solution performance, a question we will turn to next.

There were a few occasions in which problems were recalled as COMPARE 3, a problem type that was NOT used in the study. An example of this problem type is:

> Gary has 5 marbles.
>
> Joyce has 3 more marbles than Gary.
>
> How many marbles does Joyce have?

Note that this problem differs from COMPARE 4 only in one word (i.e., 'more' rather than 'less'), although changing this one word also changes the problem from a subtraction problem to an addition problem. The wording of COMPARE 3 also differs only slightly from that of COMPARE 5. In the latter, the second line reads 'He has 3 more marbles than Joyce.', while the rest of the problem is the same. Again, however, switching the names in this one line represents a dramatic change in problem structure, from a subtraction problem to an addition problem. The relation between this error and solution strategy will be discussed below.

<u>Table 5</u>

<u>Structural Recall Errors</u>

Problem Type	Recall in errors as: Easy CH1	CB1	CP4	CP3	¶	¶ Hard CH5	CB2	CP5	Total
E Change 1	-	10 (.11)			¶ ¶	1 (.01)		1 (.01)	12
A Combine 1		-			¶ ¶				0
S Compare 4	1 (.01)	6 (.07)	-	1 (.01)	¶ ¶			1 (.01)	9
Y					¶				
H Change 5	13 (.14)	1 (.01)	1 (.01)		¶ ¶	-			15
A Combine 2		16 (.18)			¶ ¶	1 (.01)	-	-	17
R Compare 5	2 (.02)	11 (.12)	32 (.36)	6 (.07)	¶ ¶				51
D					¶				
Total	16	44	33	7		2	0	2	104

Note: Numbers in parentheses are the corresponding proportions.

Solution errors and their relation to recall structural errors. A simplification bias was found in children's recall, as predicted. A more important question, however, is whether solution errors were related in any way to memory biases. To address this question, the number and types of solution errors committed by the children were scored, and the results presented in Table 6. Errors were divided into two major types, _non-conceptual errors_ and

Table 6

Solution Error Types

	Problem Type	Non-Conceptual Errors: Memory Error	Non-Conceptual Errors: Off by 1	Conceptual Errors			Other	Total
EASY	Change 1	4	1	StartSet 1			4	10
EASY	Combine 1	2	2	-----			2	6
MED	Compare 4	1	1	LargeSet 6	DifferenceSet 3	Added 3	5	19
MED	Change 5	7	5	StartSet (4)**	TransferSet 9	ResultSet 2	12	35
HARD	Combine 2	1	3	SuperSet 13	Subset 5		7	29
HARD	Compare 5	2	2	LargeSet 2	DifferenceSet 8	Added 11	4	29

Note: The maximum number of possible errors for each problem type is 90, 3 problems of each type from each of 30 children.

** These four represent occasions in which the child at first gave SOME as the answer, and then was required to give a number as the answer. These four occasions were therefore recorded, but not counted among the errors.

conceptual errors. Non-conceptual errors were ones that indicated an conceptual understanding of the problem, but a failure to execute the solution strategy properly. These errors were of two types. The first included a memory error, wherein the child performed the correct operation (addition or subtraction), but misrecalled the numbers in the problem. For example, if the problem was an addition problem involving the numbers 2 and 5, a child committing this error type might give 9 as the answer and then recall the problem with the numbers 4 and 5. The second error type involved a common bug in children's arithmetic algorithms: that of giving an answer that is off by 1. In the preceding example, a child committing this type of error would give the answer as 8 or 6.

Conceptual errors were ones that reflected a misunderstanding of the problem structure. These included giving back one of the given numbers as the answer and using the wrong arithmetic operation. Errors that could not be classified in either of these categories because it was not clear what the children had done were classified as "other errors".

The first thing to notice is that, if one uses conceptual errors as a metric for determining difficulty, the problems divide themselves into three levels. Only one conceptual error was observed for CHANGE 1 and COMBINE 1, twenty-three were observed for COMPARE 4 and CHANGE 5, and thirty-nine for COMBINE 2 and COMPARE 5. Non-conceptual errors, one the other hand, were fairly well distributed across the three levels, with CHANGE 1 and COMBINE 1 logging nine non-conceptual errors, COMPARE 4 and CHANGE 5 fourteen errors, and COMBINE 2 and COMPARE 5 eight non-conceptual errors. Clearly, what seems to determine difficulty in these problems is the ease with which their structures can be conceptualized by the child.

Comparing error types with recalled structure yielded several interesting results. First, consider COMPARE 5. The most frequent type of solution error committed here was adding the two numbers rather than subtracting them. Of the eleven observations of this error type, three were solve only problems, and eight were solve and recall problems. Of the eight problems for which there was a recall, four were recalled as COMBINE 1, two were recalled as COMPARE 3, and two were recalled as COMPARE 4. Therefore, in 75% of the occasions when the the numbers were added, the problem structure was recalled as an addition problem. The second most frequent error type was giving the cardinality of the difference set as the answer. There were eight occasions of this error, six six of which were solve and recall problems. Of these six, four of them, or 67%, were recalled as COMPARE 3, in which the difference set is owned by the person named in the final problem question (e.g., How many does Judy have?). More importantly, recall errors of these types rarely occurred in the recall only condition. When simply asked to recall COMPARE 5 problems, children misrecalled them as COMBINE 1 only 10% of the time, and they never misrecalled them as COMPARE 3. Clearly, the act of attempting a solution produced structural errors in problem representation. These results suggest that children have two default strategies for handling problems they find particularly troublesome. The first is to simplify the problem text, which sometimes results produces an errorful change in problem structure. The second is to search memory for the cardinality of the set owned by the person specified in the problem question, and return that as the answer.

These interpretations hold for COMBINE 2 problems as well. The most frequent error for this problem type was giving the cardinality of the superset as the answer. There were thirteen observations of this error, eleven of which were solve and recall problems. Of these eleven, eight were recalled as COMBINE 1 and three as COMBINE 2. Thus, in 73% of the cases where the superset cardinality was returned, the problem structure was recalled as COMBINE 1, the problem type for which this answer is appropriate. Misrecalls of this sort occurred only 23% of the time in the recall only condition, indicating a slight bias towards comprehending COMBINE 2 as COMBINE 1 that is magnified during a solution attempt. Clearly, a simple search of an incorrectly-formed problem structure was conducted in such cases, and the sought-after set cardinality retrieved.

This interpretation is further bolstered by the error patterns associated with CHANGE 5 problems. The final question to a CHANGE 5 problem requests the cardinality of the set formed 'in the beginning'. The cardinality of the start set is 'SOME', if the correct problem representation is built, but no operations performed on it. If the correct problem representation is not built because the child does not recognize the word 'SOME' to be a quantity word. In this case, the transfer set would be the first set constructed, and the cardinality of this set should be returned. The most frequent conceptual error should therefore be giving 'SOME' or the transfer set as the answer. This was precisely the case. The error of giving the transfer set as the answer was committed nine times, the most frequent conceptual error committed. On four (44%) of these occasions, the problem was recalled as a CHANGE 1 problem, as compared to only 13% of the time in the recall only condition. In the remaining two instances, the problem was recalled as a hodgepodge of CHANGE 1 and CHANGE 5, indicating severe conceptual difficulties, as follows:

> John had five marbles. (Transfer set in original problem).
>
> Susie gave him nine more marbles. (Result set in original problem).
>
> How many marbles did John have in the beginning? (Accurate final question.)

Notice that even though the problem does not make sense as an arithmetic problem, the answer given (transfer set) answers the final question appropriately Finally, on four error occasions, the child tried to give 'SOME' as the answer to the problem. (This answer was not accepted, and the children were told only numbers could be answers. They were then allowed to try again.) These results strongly indicate that children who misconceive the problem situation understand the final question as requesting a search for the first set constructed in memory, not as a request for a computation.

Notice also that several non-conceptual errors were committed on this type of problem, with the largest being a failure to recall the correct numbers. CHANGE 5 problems therefore appear to suffer two sources of difficulty. The first is a problem structure that is difficult to conceptualize, and the second is a problem length that taxes a child's memory capacity. Note that this latter problem would not necessarily be alleviated by giving the child the problem on paper. Even here, the relevant information must be held in short term memory during actual problem representation construction. The paper serves only to refresh memory for isolated bits of the problem.

A final question is how often recall successes were associated with correct solutions. Table 7 presents the numbers and proportions of correct recall of difficult problem structures following solutions. While some structural recall errors were observed, the data clearly indicate that correct solutions were overwhelmingly associated with correct structural recall.

Table 7

Correct Solutions and Correct Structural Recall

Problem Type	Recalled As:				
Compare 4	Compare 4 42 (.95)	Combine 1 2 (.05)	Not recalled 1 (.02)		Total 44 (1.00)
Change 5	Change 5f 10 (.27)	Change 5p 20 (.54)	Change 1 5 (.13)	Not recalled 2 (.05)	Total 37 (1.00)
Combine 2	Combine 2 43 (1.00)				Total 43 (1.00)
Compare 5	Compare 5 15 (.38)	Compare 4 18 (.46)	Combine 1 3 (.08)	Not recalled 3 (.08)	Total 39 (1.00)

Note: Numbers in parentheses are the corresponding proportions. Change 5f represents full correct recall of the Change 5 structure. Change 5p represents an incomplete recall, usually due to omitting parts of the third or final line in the problem.

Maximum correct in for each problem type is sixty, two solve and recall occasions from each of thirty children.

Discussion

The primary questions which this study addressed were (1) whether children spontaneously simplify difficult problems when attempting to solve them, and (2) whether these simplified memory representations affect solution success. The answer to each of these questions is unequivocally yes. Children were found to spontaneously transform difficult problem structures into simpler ones. Few transformations in the opposite direction were observed. More particularly, when problem structures were misrecalled, they tended to be recalled as simpler problems of the same type. Thus, Change 5 tended to be misrecalled as Change 1, Combine 2 as Combine 1, and Compare 5 as Compare 4. Rarely were problems misrecalled as different problem-types, and rarely were simple problems misrecalled as hard problems.

More importantly, these transformations were found to directly influence solution strategies. When children misrecalled Compare 5 (a subtraction problem) as Compare 3 (an addition problem), they tended to add the numbers in the problem rather than subtract them. Misrecalling Combine 2 as a Combine 1 problem was associated with giving the cardinality of the superset stated in the problem as the answer, an answer appropriate for Combine 1. Clearly, the solution strategies adopted by these children were driven by their understanding of the problem structures. Moreover, recall errors such as these rarely occurred in the recall only condition, indicating that a child's misunderstanding of a problem situation becomes most apparent following a solution attempt. In a sense, misrecall following a solution attempt constitutes a trace of the child's unsuccessful attempt to grasp the problem structure.

Our results also seem to imply that problem difficulty may comprise several components. For example, Compare 5 and Compare 4 problems bear identical problem structures, but differ in their linguistic forms. This linguistic difference, however, produces a 10% to 45% difference in solution success. We suggest that Compare 5's linguistic form proves troublesome for two reasons: (1) The necessity of retrieving a pronomial referent, and (2) the inclusion of a keyword (i.e., MORE) that is inconsistent with the operation to be performed (i.e., subtraction). Compare 4 requires no pronomial resolution, and its keyword (i.e., LESS) is consistent with the required subtraction operation.

The difference in solution performance between Combine 1 and Combine 2 problems suggests that other sources of difficulty exist aside from linguistic complexity. The linguistic forms used by both Combine 1 and Combine 2 are identical. What differs is their problem structures and the operations to be performed. Since children perform quite well on other subtraction problems, it is unlikely that the subtraction operation required by Combine 2 contributes significantly to its difficulty level. It appears instead that children find the problem structure itself difficult to conceptualize and manipulate. This interpretation is supported by the fact that difficult problems--and particularly Combine 2--were recalled less well and more slowly than easier problems, especially following a solution attempt. In other words, searching and retrieving information from a Combine 2 structure proved difficult for the children, particularly following a solution attempt when the textbase representation may no longer have been available.

Finally, the pattern of errors associated with Change 5 suggest two more sources of difficulty. The first is the memory capacity required by a problem in order to retain all information necessary to build an adequate structure. Change 5, unlike the other problems, requires that four lines be retained in memory rather than three, and there were more memory errors for this type of problem than for any of the others. The second source of difficulty concerns understanding the meaning of words such as 'SOME' in word problems. There are three ways such terms may be interpreted. First, SOME can be interpreted as an unknown set cardinality which must be computed. This is the correct interpretation. Second, SOME may be interpreted as an acceptable set cardinality. If 'SOME' is the cardinality of the starting set in the problem, then it constitutes an acceptable response to the query "How many __ did _ have in the beginning." Four occasions of this response were observed. Finally, 'SOME' may not be recognized as a set cardinality at all, but rather as a descriptive adjective. In this case, the first set to be created would be the transfer set, since it is the next quantity term encountered in the problem. The appropriate answer to the above query would therefore be the cardinality of the transfer set. This in fact constituted the most frequently occurring conceptual error for Change 5, a result that has been observed and similarly interpreted by other researchers (e.g., Briars & Larkin).

It is important to note that the variables suggested by our results to contribute most strongly to problem difficulty deal primarily with comprehension processes. Linguistic complexity, problem structure complexity, memory constraints, and cardinality referents all affect the ease with which one comprehends a problem story situation. Arithmetic procedural difficulties contributed very little to our observed pattern of results. This is entirely consistent with the empirical observation that children tend to perform 10%-30% worse on word problems than on comparable number problems. It is also consistent with our conjecture that comprehension skills and arithmetic skills are both involved in the development of problem-solving expertise. Indeed, in the present study, correct solutions were nearly always accompanied by correct structural recall, and errors were systematically related to structural misrecall.

These results carry important implications for theories of arithmetic problem-solving expertise development. The error types observed in the present study are ubiquitous; they have been reported by a variety of researchers, and many explanations concerning their sources have been offered. However, these explanations often rest on conjecture. Researchers have tended to focus exclusively on solution strategies and procedures, using these data to infer representations and processes that triggered them. We took a more direct approach to gaining information about the way children represent problems to themselves by requiring them to recall the problems for us. Their recall clearly showed that misunderstanding the problem structure or story situation can have disastrous results, indicating that structures cue solution strategies and procedures. Learning to carefully analyze the situation described by the story is therefore a primary step in developing one's skill in solving these types of problems.

References

Briars, D.J, & Larkin, J.H. (1984). An integrated model of skill in solving elementary word problems. Cognition and Instruction, 3, 245-296.

Carpenter, T.P., Corbitt, M.K., Kepner, H.S., Lindquist, M.M., & Reys, R.E. (1980). Solving verbal problems: Results and implications from National Assessment. Arithmetic Teacher, 28, 8-12.

Carpenter, T.P., Hiebert, J., & Moser, J.M. (1981). First grade children's initial solution procedures for simple addition and subtraction problems. Journal of Research in Mathematics Education, 12, 27-39.

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. Romberg (Eds.), Addition and subtraction: A cognitive perspective. Hillsdale, NJ: Erlbaum.

DeCorte, E., & Verschaffel, L. (1984). The influence of rewording verbal problems on children's problem representations and solutions. Paper presented at the Annual Meeting of the American Education Research Association, New Orleans, April, 1984.

Hudson, T. (1983). Correspondences and numerical differences between disjoint sets. Child development, 54, 84-90.

Keppel, G. (1973). Design and analysis: A researcher's handbook. Englewood Cliffs, NJ: Prentice-Hall.

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

Riley, M.S., Greeno, J.G., & Heller, J.I. (1983). Development of children's problem-solving ability in arithmetic. In H. Ginsberg (Ed.), The development of mathematical thinking. NY: Academic Press.

APPENDIX A

Easy Problems

Change 1.	Joe had three marbles.
	Then Tom gave him five more marbles.
1.00	How many marbles does Joe have now?
Combine 1.	Susan has four candies.
	Jane has five candies.
1.00	How many candies do they have altogether?
Compare 4.	Jack has nine puppies.
	Mary has two puppies less than Jack
.80	How many puppies does Mary have?

Hard problems

Change 5.	Vera had some dolls.
	Then Sarah gave her six more dolls.
.75	Now Vera has nine dolls.
	How many dolls did Vera have in the beginning?
Combine 2.	Mark and Sally have seven trucks altogether.
	Mark has two trucks.
.55	How many trucks does Sally have?
Compare 5.	George has eight sticks.
	He has five more sticks than Harry.
.35	How many sticks does Harry have?