Journal of Educational Psychology
1993, Vol. 85, No. 1, 7–23

0022-0663/93/$3.00

What Makes Certain Arithmetic Word Problems Involving the Comparison of Sets So Difficult for Children?

Elsbeth Stern

Arithmetic word problems with an unknown reference set, such as "John has 7 eggs. He has 4 eggs fewer [more] than Peter. How many eggs does Peter have?" are considerably more difficult for children than problems with an unknown compare set (second sentence: "Peter has 4 eggs more [fewer] than John.") Six experiments with 1st graders and kindergartners investigated reasons for this finding. Experiments 1–4 revealed that neither difficulties in processing the personal pronoun nor the use of key word strategies could explain the difficulty differences. Experiment 5 showed that most 1st graders were not aware that the difference between 2 sets can be expressed by either "In Set x there are *n* more objects than in set y" or "In Set y there are *n* fewer objects than in Set x." Experiment 6 indicated that this lack of access to flexible language use is what makes compare problems with an unknown reference set so difficult.

There is a growing interest in young children's ability to solve arithmetic word problems. Understanding and solving word problems demand the ability to access many different skills, such as language understanding, an understanding of the described situation, the ability to find an equation, and computation abilities to solve the problem. Thus, researching young children's problem-solving abilities for word problems (word problem solving) may make a valid contribution to the question of how children acquire complex problem-solving skills.

In research on young children's word problem solving, attention has turned to process models that explain why some word problems are more difficult than others. For example, there are several computer simulation models (e.g., Briars & Larkin, 1984; Dellarosa, 1986; Kintsch & Greeno, 1985; Reusser, 1989, 1990; Riley & Greeno, 1988; Riley, Greeno, & Heller, 1983) that attempt to model how text processing and mathematical knowledge are integrated to understand and solve a word problem. In research on text processing, a distinction is made between two levels of representation: The *propositional level* is based on language understanding, whereas the *situational level* characterizes the process of constructing a situational model (Kintsch, 1988; van Dijk & Kintsch, 1983) or a mental model (Johnson-Laird, 1983). In considering the extensive work on modeling the process of word problem solving cited earlier, one should recognize that different levels of constructing mental models are assumed for understanding and for solving arithmetic word problems. First, the *episodic situation model*, which has especially been emphasized by Reusser (1989, 1990), guides the understanding of specific story events, such as the temporal structure of the actions or the intentions of the actors involved. A *problem model* (Riley & Greeno, 1988) is constructed by abstracting features from the specific situation, such as names, objects, or specific intentions of the actors. It includes only the structural and relational information of the episodic problem model relevant for answering the question. The problem model guides the search for an appropriate *mathematical model*, which can either be a numerical equation or, especially for younger children, a counting strategy for addition and subtraction (Carpenter & Moser, 1983).

The problem model is derived from the activation described in the word problem and the unknown information. Simple word problems that require adding or subtracting two numbers can deal with three activities: the combination, the change, or the comparison of sets (Carpenter & Moser, 1983; Morales, Shute, & Pellegrino, 1985; Riley et al., 1983). Within these three types of word-problem groups, different kinds of word problems can be developed by varying the information that is given and the information for which one looks.

Several studies have shown that word problems demanding the same mathematical operation are not always equal in difficulty. Specifically, problems that deal with the comparison of sets are more difficult for children than other problems (Cummins, Kintsch, Reusser, & Weimer, 1988; Riley & Greeno, 1988; Riley et al., 1983; Stern, 1992; Stern & Lehrndorfer, 1992; Verschaffel, De Corte, & Pauwels, 1992). This article focuses on one well-known and often replicated difference in difficulty between two problem types that deal with the comparison of sets. Problems with

This study was presented in part in April 1991 at the annual meeting of the American Educational Research Association, Chicago.

I would like to thank Merry Bullock, Zemira Mevarech, Tina Whitley, and two anonymous reviewers for their helpful comments on an earlier version of this article, and Karin Händeler, Anne Lehrndorfer, Claudia Weber, and Heiner Welchert for their help in data collection.

Correspondence concerning this article should be addressed to Elsbeth Stern, Max-Planck-Institute for Psychological Research, Leopoldstrasse 24, 8000 Munich 40, Federal Republic of Germany. Electronic mail may be sent to Stern@mpipf-muenchen.mpg.dbp.de.

an *unknown compare set,* such as the following:

1. John has 5 marbles.
   Peter has 2 marbles fewer than John.
   How many marbles does Peter have?[1]
   (small set unknown)

or

2. John has 3 marbles.
   Peter has 2 marbles more than John.
   How many marbles does Peter have?
   (large set unknown),

are considerably easier than problems with an *unknown reference set,* such as

3. John has 5 marbles.
   He has 2 marbles more than Peter.
   How many marbles does Peter have?
   (small set unknown)

or

4. John has 3 marbles.
   He has 2 marbles fewer than Peter.
   How many marbles does Peter have?
   (large set unknown).

In Table 1 the solution rates found in different studies for elementary school children are depicted. Explaining the difference in difficulty between unknown reference set problems and unknown compare set problems was the goal of several experiments that are reported in the present study. At first glance, problems with an unknown reference set differ from problems with an unknown compare set only in the personal pronoun used in the second sentence. There is no difference in the required mathematical knowledge: To solve Problems 1 and 3, one has to subtract the numbers, and to solve Problems 2 and 4, one has to add the numbers.

Table 1
Solution Rates Reported in Studies in Which Compare Problems With Unknown Compare Set and Unknown Reference Set Were Investigated in Elementary School Children

			Proportion correct	
Study	*N*	Grade	Unknown compare set	Unknown reference set
Riley and Greeno (1988)	117	1	.22	.08
	131	2	.85	.50
	116	3	.98	.75
Morales, Shute, and Pellegrino (1985)	24	3	.52	.37
Cummins, Kintsch, Reusser, and Weimer (1988)	38	1	.57	.41
Stern (1989)	1,142	2	.85	.54
Stern (1992)	114	2	.56	.19
Stern and Lehrndorfer (1992)	45	1	.72	.37

Note. All values reflect aggregate data over addition and subtraction problems.

Briars and Larkin (1984), Riley et al. (1983), and Riley and Greeno (1988) presented models of word problem solving, called *logico-mathematical models,* that stressed the importance of mathematical knowledge. These authors' basic assumption was that mathematical knowledge develops from action-based external modeling of quantitative information to reasoning on the basis of the *quantitative part–whole schema.* Representing the part–whole schema means understanding numbers as parts of each other. The part–whole schema includes an understanding of commutativity and associativity as well as an understanding of the complementary relation of addition and subtraction (Resnick, 1989; Resnick & Greeno, 1990). According to these authors, a representation of the part–whole schema allows one not only to apply principles when performing mathematical operations but also to be aware of these principles and to be able to talk about them. Therefore, having the part–whole schema represented means being able to connect language about quantities with mathematical concepts.

According to the logico-mathematical models of word problem solving, arithmetic word problems differ in mathematical knowledge requirements. Some problems can be modeled externally and thus require only counting-based procedures, whereas other problems demand the transformation of the problem text into part–whole relations. Three levels of performance, reflecting different levels of cognitive development, were postulated by Riley and Greeno (1988) and are described here with respect to different kinds of compare problems.

At *Level 1,* children can solve compare problems with an unknown difference set, such as the following:

5. John has 5 marbles.
   Peter has 3 marbles.
   How many more marbles does John have than Peter?

These problems do not require transformation into part–whole schemata but can be solved by building a one-to-one correspondence between the two sets and counting the objects that remain (the *match–separate strategy*). Each problem-solving step can be represented externally by using objects.

Level 2 allows understanding of quantitative information that cannot be represented externally because knowledge about the relation between the sets is required to make some inferences. Compare problems with an unknown compare set can be solved on this level. For these problems it is not necessary to transform the information into a mathematical equation based on the part–whole schema. Rather, one has to represent the difference set mentioned in the second

[1] Several authors use the word *less* in these word problems, although *fewer* may be more appropriate for describing the relation between count nouns. In this article, the word *fewer* is used to describe the word problems. In German, my native language and the language in which the word problems in all experiments reported in this article were presented, no such distinction exists. Rather, the word *weniger* is used for describing the relation between both count nouns and mass nouns.

sentence as a relation between the set mentioned in the first sentence and the set asked for in the question. With this knowledge, unknown compare set problems can be solved by the *counting-all strategy* (Carpenter & Moser, 1983) if the large set is what is to be looked for and by the *separate strategy* if the small set is to be searched.

At *Level 3*, the part–whole schema is represented and can be combined with knowledge about numerical operations, such as the comparison of sets. This means knowing that three sets are involved in the comparison of sets: the *compare set*, which has to be compared to another, or *reference set*, and the difference between these two sets, the *difference set*. The comparison of sets also involves knowing that the mathematical relation between these three sets depends on whether the compare set is the *small set* and the reference set is the *large set* or the other way around. Representing the sets involved in the quantitative comparison in part–whole relations means knowing that "small set = large set − difference set" or "large set = small set + difference set" or "difference set = large set − small set." With the help of this knowledge, the textual information given in the word problem can be directly transformed into a mathematical equation. On this level, unknown reference set problems can be solved with the help of a *mathematical transformation strategy* because one only has to infer from the second sentence of the problem whether the set mentioned in the first sentence is the small set or the large set and, on the basis of this decision, either add or subtract the numbers.

According to Riley and Greeno (1988), the fact that unknown reference set problems (but not unknown compare set problems) require access to the part–whole schema is the reason for the difference in difficulty. However, Riley and Greeno's modeling of the solution of compare problems does not in all aspects fit with their own and other authors' empirical data. The model predicts that problems with an unknown difference set (such as Problem 5 cited earlier) should be easier than problems with an unknown compare set. Data reported by Riley and Greeno (1988), Cummins et al. (1988), Stern (1992), and Stern and Lehrndorfer (1992) showed that for elementary school children, problems with an unknown difference set are more difficult than problems with an unknown compare set. However, rewording compare problems with an unknown difference set (e.g., "Here are 5 birds and here are 3 worms. Suppose the birds all race over and each tries to get a worm. How many birds won't get a worm?") leads to substantial facilitation effects (Davis-Dorsey, Ross, & Morrison, 1991; Hudson, 1983). This result suggests that difficulties with text processing rather than with access to mathematical knowledge may be what makes compare problems so difficult.

A second type of word-problem-solving model (e.g., Cummins et al., 1988; Reusser, 1989, 1990) underscores the importance of text processing. According to these *text-processing models*, children's difficulties with word problems arise from a lack of textual understanding, which prevents them from making contact with relevant mathematical knowledge. For example, in the case of compare problems, children understand formulations such as "*n* more *x* than *y*" or "How many more *x* than *y*?" not as indicating relations between sets but rather as indicating simple assignments such as "There are *n x*" or "How many *x*?". To explain the difference in difficulty between unknown compare set problems and unknown reference set problems, the text-processing models assume that the language used in unknown reference set problems provides more difficulties in assigning the textual information to parts and wholes in an adequate way than does the language used in unknown compare set problems.

Performing the mathematical part–whole transformation described in Riley and Greeno's (1988) model at Level 3 is not the only way to solve unknown reference set problems. Rather, these problems can also be solved by transforming them into unknown compare set problems. To do this, the second sentence of an unknown reference set problem such as the following:

Peter has 5 marbles.
He has 2 marbles fewer than John.
How many marbles does John have?

has to be transformed into this:

John has 2 marbles more than Peter.

Performing this *linguistic restructuring strategy* allows solving of unknown reference set problems with the mathematical knowledge accessible on Level 2 of Riley and Greeno's (1988) model. According to Lewis and Mayer (1987), this transformation is the usual way of solving unknown reference set problems even for adults and is preferred because subjects attempt to match the language used to describe relational statements to the situation (the *consistency hypothesis*). Referring to the work of Huttenlocher and Strauss (1968), who showed that subjects prefer a particular ordering of information in relational statements when the spatial location of two objects is described, Lewis and Mayer (1987) argued that the information presentation in unknown compare set problems fits the preferred format better than does the information presentation in unknown reference set problems. In the second sentence of unknown compare set problems, the unknown variable is the subject, whereas in unknown reference set problems, it is the object. This makes understanding more difficult, and therefore the subjects are assumed to mentally rearrange the information such that language is consistent. However, because this transformation process is error prone, the solution rate is lower for unknown reference set problems (inconsistent language problems) than for unknown compare set problems (consistent language problems).

Because the subjects in Lewis and Mayer's (1987) study were college students, the compare problems were not the simple one-step problems described above but were more complex two-step problems, such as "At ARCO gas sells for $1.13 per gallon. Gas at Chevron is 5 cents less than gas at ARCO. How much do 5 gallons of gas cost at Chevron?" in the consistent form. In the inconsistent form, the second sentence read, "This is 5 cents more per gallon than gas at Chevron." The results showed that the probability of error was much higher with the inconsistent form. The typical error was the "reversal error," that is, the numbers were added instead of subtracted or vice versa.

Lewis and Mayer's (1987) consistency hypothesis was challenged by Verschaffel et al. (1992), who claimed that more thorough tests than those provided by Lewis and Mayer were necessary to test their hypothesis and who used eye movement procedures to obtain information about the process of problem solving. Verschaffel et al. argued that if inconsistent statements lead to confusion, subjects should spend more time reading the second sentence in inconsistent than in consistent language problems. Three experiments were conducted, two with university students and one with third graders. For one-step compare problems such as the present Problems 1, 2, 3, and 4 in the present study, more reversal errors and longer solution times were found for problems with inconsistent language (unknown reference set problems) than for problems with consistent language (unknown compare set problems) for elementary school children but not for university students. University students showed results similar to those of elementary school children only when presented with more complicated two-step problems such as those used by Lewis and Mayer (1987). Thus, under heavy demands it takes more time to solve inconsistent problems because of the longer fixation time required for the second sentence (see also Hegarty, Mayer, & Green, 1992).

Verschaffel et al. (1992) interpreted their own results as confirming Lewis and Mayer's (1987) consistency hypothesis. However, they noted that in addition to difficulties in restructuring the relational sentence in unknown reference set problems, other difficulties could be responsible for the difficulty differences between unknown compare set problems and unknown reference set problems, such as the possibility of using key word strategies successfully in unknown compare set problems or the use of the personal pronoun in unknown reference set problems. In the studies reported in this article, three factors that may account for the striking difference in difficulty between compare problems with an unknown reference set and compare problems with an unknown compare set were investigated: (a) the use of the personal pronoun, (b) the use of key word strategies, and (c) children's understanding of the symmetry of language involved in the quantitative comparison.

Use of the Personal Pronoun

According to van Dijk and Kintsch's (1983) model of text comprehension, understanding new textual information about an agent is facilitated if the new information first mentions the same agent. Therefore, the use of a personal pronoun facilitates text comprehension, at least for skilled readers. However, Verschaffel et al. (1992) showed in their Experiment 1 that for simple compare problems it took less time for university students, but not for elementary school children, to read the first two sentences of unknown reference set problems than to read the first two sentences of unknown compare set problems. In addition, it has been found in other studies that the facilitating effect from the use of personal pronouns may not occur for children (Wykes, 1981). Indeed, Oakhill and Yuill (1986) showed that children less skilled in reading comprehension have difficulties processing pronouns when complex processing is required. Under complex processing conditions, less skilled comprehenders sometimes failed to use either syntactic or semantic cues available in the text. Solving arithmetic compare word problems is a complex processing condition for first graders. Understanding relational quantitative information is a demanding process, and subjects might be overloaded when, in addition, anaphoric inferences are necessary. Thus, the use of the personal pronoun in problems with an unknown reference set may confuse children and prevent them from solving the problem. In Experiments 1–3 I tested this assumption.

Key Word Strategies

Teachers are very familiar with the fact that children can find the correct solution to a word problem without understanding it. In simple word problems that contain only two numbers that have to be either added or subtracted, the probability of guessing the correct mathematical operation is 50%. A more "elaborated" strategy of solving the problem without understanding it involves the use of key word strategies, that is, looking for verbal expressions that allude to mathematical operations. How ubiquitous the use of key word strategies may be was shown by Schoenfeld (1982). Children who were told that the word *left* used in a word problem is a hint for subtraction also subtracted when a person named "Mr. Left" was mentioned in the problem.

Compare problems may also be solved with key word strategies (Nesher & Teubal, 1975). For example, the word *more* is perceived as indicating addition, the word *fewer* as indicating subtraction. The use of key word strategies will lead to correct solutions for problems with an unknown compare set and incorrect solutions for problems with an unknown reference set. In unknown compare set problems, the language used to describe the difference between two sets and the relevant mathematical operation (addition or subtraction) are congruent; when the word *more* is used, the numbers have to be added, and when *fewer* is used, they have to be subtracted. For unknown reference set problems, the language and the mathematical operation are not congruent. Thus, if children use key word strategies to find the adequate mathematical equation, they will appear to solve problems with an unknown compare set correctly even though they may not understand them. To test whether the superior performance in solving unknown compare set problems over solving unknown reference set problems is due to key word strategy use, in Experiment 4 I included an additional indicator of word-problem understanding along with a measure of problem-solving accuracy.

Understanding the Symmetry of Language Involved in Quantitative Comparison

As discussed earlier, unknown reference set problems can be solved either by direct mathematical transformation, which requires access to part–whole representations, or by rearranging the relational information in the second sentence so that it corresponds to that in unknown compare set problems, which can be solved without access to part–whole relations (linguistic restructuring). Rearranging the relational information requires some flexibility in the use of language about the comparison of sets: One has to know that the difference between two sets can be expressed either by saying "Set 1 has *n objects fewer* than Set 2" or "Set 2 has *n objects more* than Set 1." Being aware that both sentences describe the same fact and are therefore equivalent in meaning is natural for subjects who have represented the part–whole schema. These subjects recognize that the first sentence can be expressed in the mathematical equation "Set 1 = Set 2 – *n*" and that the second sentence is equivalent to the equation "Set 2 = Set 1 + *n*," and they are aware that the equations are complementary. Whether children in beginning elementary school, who are not assumed to represent information in this formal way, nonetheless understand the symmetry of language involved in quantitative comparison necessary for the linguistic transformation was investigated in Experiment 5.

It is well known that the acquisition of linguistic expressions dealing with the comparison of sets, such as *more* and *less*, is difficult for young children (Greeno, 1989). *More* and *less* (or *fewer*) are antonyms, and the full meaning of these kinds of polar adjectives is acquired later than that of other adjectives (E. V. Clark, 1973). In addition, it has been found that young children understand the expression *more* earlier than they understand the expression *less* (Wilcox & Palermo, 1977). One explanation for this is that antonymous terms are asymmetric; the positive, or *unmarked*, term is used not only in a comparative context but also for absolute descriptions (e.g., in the case of *tall–short*, it is usual to say "He is 1.62 meters tall," but not "He is 1.62 meters short"), whereas the negative, or *marked*, term (*short*) is only understandable as a contrast to the positive unmarked term. According to H. H. Clark (1969), unmarked terms are easier to access than marked terms. With respect to the antonyms *more* and *less*, *more* is unmarked and *less* is marked, and thus it may be more difficult to solve problems containing the word *less* than those containing the word *more*.

Carey (1982) assumed that there is a simpler explanation for the later acquisition of *less* than of *more*: Children hear the word *more* in its comparative use more often than they hear the word *less*. Donaldson and Balfour (1968) showed that children ages 3–4 years have not learned that *more* and *less* are contrastive adjectives and can be used in an interchangeable way. They showed children two pictures with apple trees and told them to point either to the tree with more apples or to the tree with fewer apples. Four-year-old children always pointed to the tree with more apples. Like Carey (1982), Donaldson and Balfour (1968) explained this result in terms of the special context in which children experience the word *less*: "Don't take so many cookies, take less." Therefore children may assume that *less* means "more but not so much more." However, Klatzky, Clark, and Macken (1973) showed that the method Donaldson and Balfour (1968) used only gave the impression that the children understood *more* when in fact children used *more* and *less* to mean, respectively, *many* and *much*.

Although there are still many unresolved questions concerning the acquisition of the expressions *more* and *less*, one can undoubtedly assume that the acquisition of these words is difficult for children. However, by the time children have entered school, they normally have acquired the meanings of *more* and *less*. They pass the number conservation task (Siegler, 1981; Stern & Schneider, 1989) with the question "Which row has more?" as well as with the question "Which row has fewer?" However, children's understanding of the meaning of *more* and *less* in a qualitative context does not mean that they understand the expressions in a quantitative context, such as "*x* more [fewer] than *y*" or "How many more [fewer] *x* than *y*?" (Greeno, 1989; Resnick, 1989; Resnick & Greeno, 1990). As Greeno (1989) showed, children answered the questions "How many *x* are there?" and "Are there more *x* or fewer *y*?" correctly but failed to answer the question "How many more *x* than *y* are there?" Even if children can understand statements containing *more* and *less* [*fewer*] separately, they probably do not know their contrastive meaning. In Experiment 5 I tested whether children possess knowledge involving the symmetry of language involved in quantitative comparisons, and in Experiment 6 I tested the relation between this knowledge and performance on unknown reference set problems.

Experiments 1–3: Difficulties With Personal Pronouns

The second sentence in unknown reference set problems begins with a pronoun. To solve the problem, one has to know that this pronoun refers to the person named in the first sentence. This special language requirement appears when the numbers of objects belonging to two subjects are compared, that is, in *two-person problems*, such as Problems 1 through 4 presented earlier. In Experiments 1 and 2, use of the pronoun was avoided by repeating the name of the person mentioned in the first sentence, such as in "Peter has 5 marbles. Peter has 2 marbles fewer than John. How many marbles does John have?"

Experiment 1

Method

Subjects. Forty-two kindergarten children (23 girls, 19 boys; mean age = 5 years, 8 months; age range = 5 years, 1 month to 6 years, 5 months) from two kindergartens in Munich, Federal Republic of Germany, participated in the study.

Materials and procedure. The children were presented with four problems with an unknown reference set, two with *more* and two with *fewer*. In one of each of these problems, the name of the subject mentioned in the first sentence was repeated in the second sentence (*repeated name problems*), and in the other problem the person in the first sentence was referenced by using a pronoun (*pronoun problems*). Two compare problems with an unknown compare set (one with *more* and one with *fewer*) were also presented. These six problems were presented with 14 other two-step word problems that are not discussed here. The numbers used in the problems were chosen such that their sums were not greater than 8.

The children were tested individually by a female research assistant in a separate room of the kindergarten in two sessions separated by 1 week. The order of problem presentation was randomized for each child. The problems were read out loud once to each child, and if the child asked for repetitions, they were read a second time. The child had to answer each problem by stating a number, which was recorded by the experimenter.

Results and Discussion

A problem was evaluated as correctly solved if the child answered with the correct number. Following Cummins et al. (1988), I classified incorrect answers as given-number errors or wrong-operation errors. The wrong-operation error type, called "reversal error" by Lewis and Mayer (1987), is characterized by adding the numbers instead of subtracting them, or vice versa. According to data from Cummins et al. (1988), this error is associated with changing the problem model of the problem. Stern and Lehrndorfer (1992) have shown that a wrong-operation error is typical for compare problems with an unknown reference set. In several studies (Cummins et al., 1988; De Corte & Verschaffel, 1985; Riley et al., 1983), it has been shown that a frequent wrong answer involves repeating one of the numbers given in the problem. Cummins et al. (1988) showed that these given-number errors are associated with a failure to understand the gist of the problem.

Answers were assigned to the following categories: (a) correct number, (b) given-number error, (c) wrong-operation error, (d) other number error, and (e) no number answer or no answer at all. Table 2 depicts the solution rates for the compare problems.

Whether the small set or the large set of a problem was unknown was not relevant for the question addressed in this experiment. Therefore, the dependent variable for statistical analysis was constructed by averaging the number of correctly solved problems for each problem type (unknown compare set: $M = .30$, $SD = .38$; unknown reference set, pronoun: $M = .27$, $SD = .39$; unknown reference set, repeated name: $M = .26$, $SD = .34$). A repeated measures analysis of variance (ANOVA) with the within-subject factor of problem type (unknown compare set, unknown reference set with pronoun, unknown reference set with repeated name) revealed no significant results, $F(2, 82) = .15$, $MS_e = .09$. Avoiding use of the pronoun did not make unknown reference set problems easier for kindergarten children. However, the low solution rates for unknown compare set problems and the high rates of given-number errors in all problem types indicate that only a small proportion of the children were able to understand quantitative relational statements with *more* or *fewer* at all, and thus repeating the name rather than using a pronoun probably would not have had an effect for such young children.

Because the results of most of the studies mentioned in Table 1 show that the majority of first graders can solve unknown compare set problems, one might expect this age group to profit from avoiding the pronoun in unknown reference set problems. Therefore, in a second experiment, the problems considered in Experiment 1 were presented to first graders.

Experiment 2

Method

Subjects. Forty-eight first graders (28 boys, 20 girls; mean age = 7 years, 7 months) from two different classes of an elementary school in the Munich area of the Federal Republic of Germany participated in the study. The test was administered during the last month of the school year.

Materials and procedure. The two unknown compare set problems and the four unknown reference set problems were similar to those used in Experiment 1 except that the numbers used were between 2 and 9. The problems were presented together with 16 other one-step and more-step word problems (not discussed here) in a group test I conducted in each class. Each child received an exercise book with 2–4 word problems on each page. The children were instructed (a) to read along as the experimenter read each problem out loud; then (b) to write down the answer (only the number) in a square located to the right of the question; and finally (c) to write down the mathematical equation on a line printed below the question. After 90 s the experimenter started to read the next problem. Two orders of problem presentation were prepared, one for each class.

Table 2
Solution Rates and Rates of Errors for the Compare Problems in Experiment 1

Solution	Unknown compare set: Large set unknown	Unknown compare set: Small set unknown	Unknown reference set, pronoun: Large set unknown	Unknown reference set, pronoun: Small set unknown	Unknown reference set, repeated name: Large set unknown	Unknown reference set, repeated name: Small set unknown
Correct	.26	.33	.29	.26	.29	.24
Given-number error	.45	.41	.41	.41	.33	.36
Wrong-operation error	.10	.05	.05	.02	.07	.14
Other-number error	.19	.17	.19	.26	.19	.19
No number answer	.00	.05	.07	.05	.12	.07

Results and Discussion

The children's solutions were classified into the following seven categories:

1. Completely correct: The number in the square was correct and the equation was appropriate.
2. Only answer correct: The number in the square was correct but no or an inappropriate equation was written.
3. Arithmetic error: The equation was appropriate but the answer deviated by one from the correct answer.
4. Wrong-operation error: An inappropriate equation was chosen, and the number in the square was the result of the inappropriate mathematical operation.
5. Given-number error: The number in the square was a number mentioned in the problem. The equation was not considered in this category.
6. Other: For example, an inappropriate operation was stated but the number in the square was not the correct result of this equation; a number was written in the square that was neither the correct answer nor a number mentioned in the problem; or an equation was stated but no number was written into the square.
7. No answer: Nothing was written down.

It has been shown (e.g., Jaspers, 1991) that often when students do not understand word problems, they use a strategy of simply adding the numbers mentioned. For the kindergarten children in Experiment 1, no such response set was found. It seems that children develop such "survival" strategies to compensate for deficiencies only after they enter elementary school. Seven subjects of Experiment 2 showed this response set: They added the numbers for all 6 compare problems discussed here and they also added the numbers for at least 11 of the 16 additional problems, although only 5 of these required addition. The data of these children were excluded because considering them would lead to an overestimation of the solution rates for all compare problems with the large set unknown. The answers from the remaining 41 children were used in the analyses.

Table 3 depicts the frequencies of each category for each problem. Whether the small set or the large set of a problem was unknown was not relevant for the question addressed in this experiment. Therefore, the dependent variable for statistical analysis was constructed by averaging the number of problems correctly classified (unknown compare set problems: M = .51, SD = .44; unknown reference set problems with pronoun: M = .27, SD = .32; unknown reference set problems with repeated name: M = .26, SD = .36). A repeated measures ANOVA with the factor of problem type (unknown compare set, unknown reference set with pronoun, unknown reference set with repeated name) revealed a significant effect, $F(2, 80) = 9.61$, $p < .001$, $MS_e = .09$. Post hoc Tukey tests indicated that performance on unknown compare set problems was better than on both types of unknown reference set problems ($p < .05$), whereas repeated name problems and pronoun problems did not differ in solution rate. Similar results were found with the same ANOVA when not only "completely correct" answers but also "only answer correct" answers (Category 2) and "correct but arithmetic error" answers (Category 3) were included (unknown compare set problems: M = .60, SD = .44; unknown reference set problems with pronoun: M = .35, SD = .36; unknown reference set problems with repeated name: M = .38, SD = .42), $F(2, 80) = 6.95$, $p < .01$, $MS_e = .11$. Post hoc Tukey tests indicated that performance on unknown compare set problems was better than on both types of unknown reference set problems ($p < .05$), whereas repeated name problems and pronoun problems did not differ in solution rate. Therefore, avoiding the pronoun did not make unknown reference set problems easier for first graders.

I also analyzed whether different problem types differed in the frequency of given-number errors and wrong-operation errors by conducting two repeated measures ANOVAs with the factor of problem type and with averaged error frequencies as dependent variables. For wrong-operation errors (unknown compare set problems: M = .04, SD = .13; unknown reference set problems with pronoun: M = .28, SD = .30; unknown reference set problems with repeated name: M = .32, SD = .40), the analysis revealed a significant effect, $F(2, 80) = 11.87$, $p < .001$, $MS_e = .08$. Post hoc Tukey tests indicated that both types of unknown reference set problems were more prone to wrong-operation errors than were unknown compare set problems ($p < .05$), whereas repeated name problems and pronoun problems did not differ. For given-number errors (unknown compare set problems: M = .32, SD = .38; unknown reference set problems with pronoun: M = .32, SD = .38; unknown reference set problems

Table 3
Solution Rates and Rates of Errors for the Compare Problems in Experiment 2

	Unknown compare set		Unknown reference set, pronoun		Unknown reference set, repeated name	
Solution	Large set unknown	Small set unknown	Large set unknown	Small set unknown	Large set unknown	Small set unknown
Completely correct	.44	.59	.20	.34	.29	.22
Only answer	.00	.02	.07	.02	.02	.07
Arithmetic error	.10	.05	.07	.00	.07	.07
Wrong-operation error	.02	.05	.27	.29	.32	.32
Given-number error	.37	.27	.29	.34	.29	.27
Other	.05	.02	.05	.00	.00	.00
No answer	.00	.00	.05	.00	.00	.05

with repeated name: $M = .28$, $SD = .43$), the ANOVA revealed no significant effect, $F(2, 80) = .37$, $MS_e = .05$.

Experiment 3

One might argue that repeating the name instead of using a pronoun in Experiments 1 and 2 was artificial and unfamiliar and therefore could have been confusing to children, leading them to assume that another person also named "Peter" was referred to second sentence. Therefore, to avoid the pronoun, I developed problems in which two different sets of objects were compared without reference to a person's ownership. To investigate the role of the personal pronoun, I compared the solution rate of these *two-set problems* with the solution rate of two-person problems.

Method

Subjects. Forty 1st graders (24 girls, 16 boys; mean age = 7 years, 6 months) from two different after-school centers in Munich, Federal Republic of Germany, participated in the experiment.

Materials. The two-set problems were stated as follows[2]:

Unknown compare set problems:

6. There are 5 lions in the zoo.
   There are 2 more tigers than lions in the zoo.
   How many tigers are in the zoo?
   (large set unknown)
7. There are 3 lions in the zoo.
   There are 2 fewer tigers than lions in the zoo.
   How many tigers are in the zoo?
   (small set unknown)

Unknown reference set problems:

8. There are 5 lions in the zoo.
   There are 2 fewer lions than tigers in the zoo.
   How many tigers are in the zoo?
   (large set unknown)
9. There are 5 lions in the zoo.
   There are 2 more lions than tigers in the zoo.
   How many tigers are in the zoo?
   (small set unknown)

Procedure. Subjects were asked to solve eight compare problems: four problems with an unknown compare set and four problems with an unknown reference set. Two of the problems of each type were two-person problems, and two of the problems of each type were two-set problems. In half of the problems, the large set was known and the small set was looked for, and in the other half it was the other way around. In the two-person problems, one actor was female and the other was male. Four number triples were chosen for use in the problems: 2, 4, 6; 3, 5, 8; 3, 6, 9; and 2, 5, 7. These triples were chosen to make the distinction between arithmetic errors and given-number errors possible. Arithmetic errors occur if something goes wrong when using an arithmetic strategy, and frequently the answer given deviates by one from the correct answer. Therefore, to distinguish arithmetic errors from given-number errors, the problems were designed so that the correct answer plus one or minus one could not be identical to any of the numbers given in the problem.

The children were tested individually by a female research assistant in a separate room at the after-school centers. Order of problem presentation was randomized for each child. The problems were presented in written form on an index card and, in addition, were read out loud once to each child. Children were asked not just to say the correct number but to state the answer in a whole sentence—for example, "There are 7 tigers in the zoo." If a child nonetheless answered only by saying a number—for example, "seven," the instructor said, "What do you mean by seven? Please use a sentence." If the child still stated only a number on this second turn, the answer was scored as incorrect. Children also had to tell how they worked out the answer, for example, "5 plus 2." The experimenter wrote down the children's answers.

Results and Discussion

To evaluate whether a problem was solved correctly, the following three criteria were considered:

1. Did the child answer the given question or another one? For the two-set problems, the child's sentence was scored as correct if the child answered "5 tigers are in the zoo," "5 tigers," or "5 of them." The child's sentence was scored as incorrect if, for example, the child said *lions* instead of *tigers* or mentioned the word *more* or *fewer* in the answer. For the two-person problems the sentence was scored as correct if the child answered "Peter has 5 marbles" or "He has 5 marbles." The sentence was scored as incorrect if the child used the wrong name or the wrong gender pronoun or mentioned *more* or *fewer*.
2. Was the correct mathematical operation (addition or subtraction) chosen?
3. Was the correct number mentioned in the child's sentence?[3]

[2] The formulation in these word problems sounds a little strange in English. The expression "there are" can be translated into German either by "es sind," or "es gibt." Using "es sind" in the context of these problems would sound strange in German, too, whereas "es gibt" (verbatim translation is "it gives," which does not make sense in English in the context of the word problems) is an appropriate formulation. The verbatim formulation of the problems was "Im Zoo gibt es 5 Tiger. Es gibt 3 Tiger mehr als Löwen. Wie viele Löwen gibt es?"

[3] In Experiment 2 some subjects were excluded from the statistical analysis because they always added the two numbers used in the word problem. In contrast, in Experiment 3 and in all subsequent experiments on word problem solving presented in this article (Experiments 4 and 6), no subjects were excluded, because this response set would not have altered the results. Two reasons may be responsible for the difference between Experiment 2 and the other experiments: First, in Experiment 2 the data were collected in a group setting, and the children thus knew that the experimenter could not control whether they tried to understand the word problem or whether they only wrote down the result obtained by the use of a surface strategy. All other experiments were conducted in individual sessions, and this direct personal contact between the subjects and the experimenter might have encouraged the children to take the task more seriously. Second, in the experiments conducted in individual sessions, children were

Table 4
Solution Rates and Rates of Errors for the Compare Problems in Experiment 3

	Unknown compare set				Unknown reference set			
	Two-person problems		Two-set problems		Two-person problems		Two-set problems	
Solution	Large set unknown	Small set unknown	Large set unknown	Small set unknown	Large set unknown	Small set unknown	Large set unknown	Small set unknown
Completely correct	.43	.45	.55	.48	.33	.25	.20	.18
Correct but arithmetic error	.00	.03	.03	.00	.00	.05	.00	.03
Wrong-operation error	.05	.05	.08	.05	.23	.20	.30	.35
Given-number error	.13	.10	.13	.08	.10	.15	.13	.05
Other	.40	.38	.23	.40	.35	.35	.38	.40

A problem was evaluated as "completely correctly" solved if all three criteria were fulfilled. If the correct mathematical equation was chosen and the child's sentence was formulated in an adequate way but the number mentioned in the child's sentence deviated by one from the correct number, it was assumed that the wrong number was caused by an arithmetic error and not by misunderstanding the structure of the problem; therefore the answer was coded as "correct but arithmetic error."

An incorrectly solved problem was classified as a wrong-operation error if the chosen mathematical operation and the number worked out were wrong but the child's sentence was appropriate for the question. An answer was also classified as a wrong-operation error if the result of the wrong mathematical operation deviated by one from the correct result, because it was assumed to be an arithmetic error. An example of a wrong-operation error was answering Problem 9 with "7 tigers are in the zoo." If the wrong operation was chosen but the answer did not make any sense (e.g., "7 more tigers are in the zoo"), the answer was coded in the category "other." A given-number error was coded if a number of the problem was mentioned in an adequate sentence. The mathematical operation (addition or subtraction) stated by the children was not considered in this evaluation because a given-number error is not the result of a mathematical operation, and therefore children either did not provide a mathematical operation at all or they seemed to guess at one.

The proportions of correctly and incorrectly solved problems are depicted in Table 4. Whether the small set or the large set of a problem was unknown was not relevant for the question addressed in this experiment, and therefore the dependent variable for statistical analysis was constructed by averaging the number of correctly classified problems for two-person unknown compare set problems (M = .44, SD = .38), for two-set unknown compare set problems (M = .51, SD = .45), for two-person unknown reference set problems (M = .29, SD = .42), and for two-set unknown compare set problems (M = .19, SD = .39). A 2 × 2 repeated measures ANOVA with the factors of problem type (unknown compare set problems vs. unknown reference set problems) and problem content (two-person problems vs. two-set problems) showed a main effect for problem type, $F(1, 39) = 15.66$, $p < .001$, $MS_e = .14$, no main effect for problem content, $F(1, 39) = .08$, $MS_e = .08$, and only a weak interaction, $F(1, 39) = 3.89$, $p < .06$, $MS_e = .08$. The same analysis was conducted by including the "correct but with arithmetic error" results (two-person unknown compare set problems: M = .45, SD = .39; two-set unknown compare set problems: M = .53, SD = .45; two-person unknown reference set problems: M = .31, SD = .45; two-set unknown compare set problems: M = .20, SD = .41). The 2 × 2 repeated measures ANOVA showed a main effect for problem type, $F(1, 39) = 13.25$, $p < .001$, $MS_e = .16$, no main effect for problem content, $F(1, 39) = .21$, $MS_e = .07$, and an interaction, $F(1, 39) = 4.44$, $p < .05$, $MS_e = .08$. Simple effects tests showed that performance was worse on problems with an unknown reference set than on problems with an unknown compare set and that within unknown reference set problems, performance was worse on two-set problems than on two-person problems ($p < .05$). Overall, the data fail to show that use of pronouns in the second sentence is a source of children's difficulty with unknown reference set problems.

I also analyzed whether different problem types differed in the frequency of given-number errors and wrong-operation errors by conducting a 2 × 2 repeated measures ANOVA with the factors of problem type and problem content and with averaged error frequencies as dependent variables. For wrong-operation errors (two-person unknown compare set problems: M = .05, SD = .22; two-set unknown compare set problems: M = .06, SD = .23; two-person unknown reference set problems: M = .21, SD = .37; two-set unknown compare set problems: M = .33, SD = .43), the ANOVA showed a main effect for problem type, $F(1, 39) = 12.94$, $p < .001$, $MS_e = .14$, whereas the other effects were not significant at the .05 level. For given-number errors (two-person unknown compare set problems: M = .11, SD = .29; two-set unknown compare set problems: M = .10, SD = .28; two-person unknown reference set problems: M = .13, SD = .29; two-set unknown compare set problems: M = .09, SD = .22), the ANOVA showed no significant effects at the .05 level.

asked to state an answer sentence, whereas in Experiment 2 they only had to write down a number and an equation. Subjects who only added the two numbers would hardly be able to state an appropriate answer sentence and therefore would not fulfill the criteria for solving the problem correctly.

As predicted, problems with an unknown reference set but not problems with an unknown compare set were prone to wrong-operation errors. From Table 4 it can be computed that the mean solution rate ("correct but arithmetic error" included) of unknown compare set problems (.49) is similar to the sum of the mean solution rate and the mean rate of wrong-operation errors for unknown reference set problems (.52).

Experiment 4: Use of Key Word Strategies

Being able to solve a word problem, especially if it is a problem with only two numbers, does not necessarily mean that the gist of the problem has been understood or, in other words, that the appropriate problem model has been built. The use of key word strategies—that is, adding the numbers when *more* is used in a word problem and subtracting when *fewer* is used—will lead to correct solution of problems with an unknown compare set and incorrect solution of problems with an unknown reference set. With the help of key word strategies, children can solve unknown compare set problems without understanding them. To test whether children's good performance in compare problems with an unknown compare set occurs because they use key word strategies, one must ask whether the children who correctly solve the problem also understand it.

Additional evidence for building an appropriate problem model is provided by asking a child to *retell* the problem (Cummins et al., 1988; De Corte & Verschaffel, 1985, 1987). For most problems, verbatim storage of a problem exceeds working memory capacity. Therefore, to store a problem requires comprehension; in other words, one has to transform the information into a problem model (Reusser, 1989; Riley & Greeno, 1988) that is less capacity-demanding than the verbatim verbal information.

Although there are good reasons to consider the ability to retell a word problem as an indicator of understanding, there is not a one-to-one correspondence between understanding and being able to retell a problem. De Corte and Verschaffel (1987) provided some evidence that children can sometimes retell a problem they cannot solve because they lack an appropriate internal representation of the problem's essential elements and relations; alternatively, sometimes children who have understood a problem make mistakes in retelling it, for example, by forgetting to mention some information. Also, the relation between retelling and understanding is affected by the fact that some word problems are longer than others. For example, change problems with an unknown change set ("Peter had 5 marbles. Then John gave him some other marbles. Now Peter has 8 marbles. How many marbles did John give to Peter?") have four sentences, whereas compare problems have only three sentences. Therefore, retelling a change problem may present more difficulties than retelling a compare problem, although compare problems present more difficulties in solving. However, because unknown compare set problems and unknown reference set problems have exactly the same number of words, it is reasonable to explain differences in retelling performance between these problem types by difficulties in understanding the gist of the problem.

I hypothesized that if first graders solved problems with an unknown compare set better than problems with an unknown reference set because they used key word strategies yet did not really understand the problems, their performance in retelling for the two problem types would be equivalent.

Method

Subjects. Twenty-nine 1st graders from two different after-school centers in Munich, Federal Republic of Germany (10 boys, 19 girls; mean age = 7 years, 6 months) participated in the study.

Materials and procedure. Subjects were asked to retell four problems with an unknown reference set and four problems with an unknown compare set. Half of the problems had an unknown large set and thus the adequate mathematical operation was addition; half of the problems had an unknown small set and thus the adequate mathematical operation was subtraction. All problems were two-person problems such as Problems 1–4. The children were tested individually in a separate room of the after-school center by a male research assistant. The problems were spoken out loud to the children from a tape recorder. The children's retelling was taped.

To demonstrate that the unknown compare set problems were solved more often than the unknown reference set problems from this sample and to find out whether the correspondence between retelling errors and solution errors was similar to the results of Cummins et al. (1988), I also asked that the children solve the problems. After the children had retold the story, it was read again by the experimenter; then the children had to solve the problem by providing a sentence and telling the mathematical operation, as in Experiment 3.

To familiarize children with the procedure, the experimenter began the session with two practice items (unknown combine set problems from Riley et al., 1983). To avoid confusion that might arise from hearing the language typical for compare problems too repetitively, the experimenter presented a filler item (change problem from Riley et al., 1983) after every two compare items. After half of the problems had been presented, there was a break of about one-half hour. The names, objects, and numbers used were different in every problem. All numbers were chosen such that their sums were less than 10. Two orders were prepared: In one, no two critical problems followed successively; the second order was the reverse of the first.

The word-problem retellings were classified into the following five categories:

1. Correct recall: verbatim or nearly verbatim.
2. Correct gist: The gist and language of the problem were maintained but the details were not correct; names, objects, or numbers were changed.
3. The gist of the problem was maintained, but the language was changed, as, for example, by replacing "He [John] has 4 marbles fewer than Peter" with "Peter has 4 marbles more than John."
4. Retelling another meaningful problem: (a) An unknown reference set problem was retold as an unknown compare set problem, or vice versa, by changing the gist. For example, "He [John] has four marbles fewer than Peter" was replaced by "Peter has 4 marbles fewer than John." (b) Another problem was retold, for

example, a compare problem with an unknown difference set such as Problem 5 or a combine problem such as "John has 5 marbles. Peter has 3 marbles. How many marbles do John and Peter have altogether?"

5. Nonsense retelling: Some critical information was omitted, the question did not fit the story, or there was no meaningful relation between the sets, as, for example, in "Peter has 3 marbles. Peter has 5 marbles more than John. How many more marbles does Peter have than John?"

The solutions of the problems were analyzed in the same way as in Experiment 3.

Results and Discussion

In Table 5 the numbers of problems in each retelling and solution category are depicted. There was a close relation between solving a problem correctly and retelling it accurately. Furthermore, if a problem with an unknown reference set was retold as a problem with an unknown compare set, it was prone to being a wrong-operation error.

For statistical analyses, a retelling score was built by giving one point for each retelling in Categories 1–3 (unknown compare set problems: $M = 2.6$, $SD = 1.6$; unknown reference set problems: $M = 1.6$, $SD = 1.5$), and a solution score was built as in Experiment 3 (unknown compare set problems: $M = 2.8$, $SD = 1.0$; unknown reference set problems: $M = 1.4$, $SD = 1.6$). A 2 (problem type: unknown compare set vs. unknown reference set) × 2 (task: retelling vs. solving) ANOVA with repeated measures on both variables was performed on the dependent variable of number of problems performed (solved or retold) correctly. Only the main effect of problem type was significant, $F(1, 28) = 23.3$, $p < .001$, $MS_e = 1.81$. Children were much better at performing (solving or retelling) problems with an unknown compare set than problems with an unknown reference set. If one assumes that correct retelling is an indicator of understanding a word problem, most children at the end of the first grade understand problems with an unknown compare set. Thus, there is no support for the claim that children are better at solving problems with an unknown compare set than problems with an unknown reference set just because they use key word strategies.

Experiment 5: Understanding the Symmetry of Language Involving Quantitative Comparisons

When children understand that the statement "There are *n* more *x* than *y*" is the same as "There are *n* fewer *y* than *x*," they understand the symmetry of comparison. This knowledge is necessary to solve unknown reference set problems by transforming them into unknown compare set problems. In Experiment 5 I tested whether first graders possessed this knowledge by asking whether they were able to recognize that both statements meant the same thing. Children received a picture with two sets of objects and were asked whether verbal statements about the difference between these two sets were valid or not. In each task, the children

Table 5
Number of Correct and Incorrect Solutions and Retellings in Experiment 4

Retelling category	Solution: Correct	Arithmetic error	Wrong-operation error	Given-number error	Other	Total
Unknown compare set						
Retelling correct						
Verbatim	50	2	1	3	0	56
Details wrong	0	2	1	1	3	7
Language changed	4	0	4	0	7	15
Retelling incorrect						
Unknown reference set problem	2	0	0	0	0	2
Other problem	10	2	4	2	5	23
Nonsense problem	8	2	1	2	0	13
Total	74	8	11	8	15	116
Unknown reference set						
Retelling correct						
Verbatim	18	1	4	0	0	23
Details wrong	3	0	0	0	1	4
Language changed	5	1	7	4	4	21
Retelling incorrect						
Unknown compare set problem	6	1	20	5	1	33
Other problem	5	0	3	3	5	16
Nonsense problem	2	0	8	7	2	19
Total	39	3	42	19	13	116

received two statements (a statement pair) about the difference between the two sets, such as "There are 3 more *x* than *y*" and "There are 3 fewer *y* than *x*," and they had to say whether both were incorrect, both were correct, or one was incorrect and the other was correct. If children did not know that the difference between two sets could be described by using both the expressions *more than* and *fewer than*, their performance would be lowest when both statements were correct, because they would not understand that the two statements were equivalent in meaning and would agree to only one of the statements even though both were correct.

Because I hypothesized that children have particular difficulties with understanding the *symmetry* of relational quantitative statements containing "*n* more than" and "*n* fewer than," but not with understanding the single statements, I expected performance to be better if one statement was correct and one was incorrect than if both statements were correct. There are different ways to design statement pairs with one incorrect and one correct statement: Both statements can have the same relational term, for example, "There are *n* more *x* than *y*" and "There are *n* more *y* than *x*." Another possibility is that one statement can contain *more* and the other *fewer*, such as "There are *n* more *x* than *y*" and "There are *n* fewer *x* than *y*." Under both conditions, to perform correctly, it is only necessary to understand the expressions "*n* more than" and "*n* fewer than," but it is not necessary to understand the symmetry of these expressions.

I also expected children to perform better on tasks in which both statements were wrong than on tasks in which both statements were correct. Children who do not understand the symmetry of language involved in quantitative comparison might have developed rules such as "If a sentence using '*n* more than' describes the difference between two sets correctly, this difference can't be described by a sentence using '*n* fewer than'." However, this does not mean that children have developed a rule such as "If a sentence using '*n* more than' does not describe the difference between two sets correctly, a sentence with '*n* fewer than' must be the adequate description of the difference." The first rule does not imply the second one. Children who assume that there is only one correct way to describe a difference between two sets can be aware of the fact that there are several possibilities to describe the difference between two sets in an incorrect way. Therefore, one can find out that two statements about the quantitative comparison are incorrect without having understood the symmetry of language involved in the quantitative comparison.

In school, children often experience the situation in which two answers are presented and only one is correct; thus, they may hesitate to say that both of the answers are correct. To control for this possibility, statement pairs referring to the whole sets were included. These statement pairs were either both incorrect, both correct, or one was incorrect and one was correct. Because first graders are expected to understand quantitative assignments, response sets have to be assumed if performance under the "whole set, both correct" condition is worse than that under the "whole set, both incorrect" and the "whole set, one incorrect and one correct" conditions.

Thus, a single statement either referred to a difference set (D) or a whole set (W) and was correct (c) or incorrect (i). If the statement referred to the difference set, either the expression *more* (m) or the expression *fewer* (f) was used. To control for children's tendencies to agree with the first statement of each pair, the order of the correct statement (first vs. second) was varied under the "one incorrect and one correct" conditions. Altogether, 14 statement pairs were designed: Wi–Wi, Wc–Wc, Wi–Wc, Wc–Wi, Dim–Dcm, Dim–Dcf, Dif–Dcm, Dif–Dcf, Dcm–Dif, Dcf–Dim, Dim–Dif, Dif–Dim, Dcm–Dcf, and Dcf–Dcm.

The main hypothesis of this study was that performance on the Dcm–Dcf (and Dcf–Dcm) task would be worse than performance on all other tasks. However, the reason for poor performance on these two tasks might be that subjects hesitate to say that both statements are correct. Therefore, one can assume that difficulties with the Dcm–Dcf (and Dcf–Dcm) task are due to difficulties in understanding the symmetry of language involved in quantitative comparison only if performance on the Wc–Wc task is not worse than performance on the Wi–Wi task and the Wi–Wc (and the Wc–Wi) task.

Method

Subjects. Forty-seven 1st graders (24 boys, 23 girls; mean age = 7 years, 3 months) from after-school centers in Munich, Federal Republic of Germany, participated in the study.

Materials. The children were presented a short story about two enterprising siblings, Peter and Susan. For example, one story began, "Today, Peter and Susan are visiting a farm to look at all the animals." Then the children were shown a picture with 6 cows and 4 pigs and were presented trials such as this one:

> Susan says: "There are 2 more cows than pigs." Peter says: "There are 2 fewer pigs than cows." Who is right, nobody, Susan, Peter, or both? (Dcm–Dcf)

In this case, both statements refer to the difference set—the first by using *more*, the second by using *fewer*—and both statements are correct. The statement pairs under the "both incorrect" condition were as follows:

> Susan says: "There are 2 fewer cows than pigs." Peter says: "There are 2 more pigs than cows." Who is right, nobody, Susan, Peter, or both? (Dif–Dim)

An example of the "one correct and one incorrect" statement pairs is as follows:

> Susan says: "There are 2 fewer cows than pigs." Peter says: "There are 2 fewer pigs than cows." Who is right, nobody, Susan, Peter, or both? (Dif–Dcf)

In addition, some stories contained correct statement pairs that referred to the whole set,

> Susan says: "There are 6 cows." Peter says: "There are 4 pigs." Who is right, nobody, Susan, Peter, or both? (Wc–Wc),

incorrect statement pairs that referred to the whole set,

> Susan says: "There are 4 cows." Peter says: "There are 6 pigs." Who is right, nobody, Susan, Peter, or both? (Wi–Wi),

and statement pairs with one statement correct and one statement incorrect that referred to the whole set,

> Susan says: "There are 5 cows." Peter says: "There are 4 pigs." Who is right, nobody, Susan, Peter or both? (Wi–Wc).

Procedure. The children were tested individually in a separate room at the center by a male research assistant. The text (story and statement pairs) was read to the child and also presented in a written form so that the child could read along. The children were to answer each question immediately.

The children were presented two stories: the story described in the preceding examples and a second story about two children who are going to the zoo and are watching the animals. The order of the two stories was random. For every picture story, the 14 statement pairs mentioned earlier were presented. Four sessions, which were completed within 2 weeks, were conducted with every child: two sessions for each picture story. At the beginning of each session, the picture story was presented and then seven statement pairs were presented successively. The order of the four possible answers was randomized between the statement pairs, and the order of the statement pairs was mixed randomly for every child.

Results and Discussion

The data of 6 subjects were excluded from the experiment because they answered all statement pairs concerning the difference sets with "both incorrect." I assumed that this was the effect of a response set. Not excluding these children would have led to an overestimation of the performance for the "both incorrect" statement pairs and to an underestimation of the performance for all other statement pairs. No response set was found for the other answer possibilities. Therefore, the data of 41 subjects were retained for statistical analyses.

The frequency of correct answers is depicted in Table 6. The data show that, as predicted, most difference-set mistakes occurred under the "both correct" condition. For statistical analyses, scale scores for each subject were constructed by averaging across the two equivalent statements for each set type: difference set, both correct (Dcm–Dcf, Dcf–Dcm); difference set, both incorrect (Dim–Dif, Dif–Dim); difference set, first correct, second incorrect (Dim–Dcf, Dcf–Dim, Dim–Dcm, Dif–Dcf); difference set, first correct, second incorrect (Dcm–Dif, Dcf–Dim); whole set, both correct (Wc–Wc); whole set, both incorrect (Wi–Wi); and whole set, one correct, one incorrect (Wi–Wc, Wc–Wi). A repeated measures ANOVA with the factor of scale produced a significant result, $F(6, 240) = 24.69$, $p < .001$, $MS_e = .07$. The hypothesis that performance was worse under the "difference set, both correct" condition than under all other conditions was confirmed by a post hoc Tukey test ($p < .05$). Tukey's test also indicated that performance in the three whole set conditions was better than in all difference set conditions ($p < .05$).

Poor performance under the "difference set, both correct" condition cannot be explained by a reluctance to say that both statements are correct because under the "whole set,

Table 6
Proportions of Correct Answers in Experiment 5 (Aggregated Over the Two Picture Stories)

Statement pairs	Proportion correct
Whole set	
Both incorrect	.84
Both correct	.93
One incorrect, one correct	
First incorrect, second correct	.94
First correct, second incorrect	.84
M	.89
Difference set	
First incorrect, second correct	
Both "more"	.53
First "more," second "fewer"	.64
First "fewer," second "more"	.58
Both "fewer"	.54
M	.57
First correct, second incorrect	
First "more," second "fewer"	.65
First "fewer," second "more"	.62
M	.63
Both incorrect	
First "more," second "fewer"	.68
First "fewer," second "more"	.61
M	.65
Both correct	
First "more," second "fewer"	.38
First "fewer," second "more"	.34
M	.36

both correct" condition, nearly all children performed correctly.

The solution rates for the "difference set, both incorrect" and "difference set, one correct, one incorrect" statement pairs are both about .60. Although this is clearly higher than chance (.25), it indicates that children have some difficulties with solving these problems too. Probably, despite dividing the tasks into four sessions, children sometimes got confused because they heard *more than* or *fewer than* too often and therefore made some errors.

The results indicate that although the majority of the children could understand each statement of the form "*x* has *n* more objects than *y*" and "*y* has *n* objects fewer than *x*," they did not realize that these statements described the same event, that is, that the statements represented different sides of the same coin. If children accepted one verbal statement concerning the difference of two quantities as true, they appeared not to access other verbal knowledge that described the same situation in different words. The same children who were unaware of the fact that a difference between sets could be described with either *more* or *fewer* could understand that two statements, one with *more* and one with *fewer*, were incorrect, as the better performance under the "difference set, both incorrect" condition than under the "difference set, both correct" condition showed.

Without knowing that one can use either *more* or *fewer* to describe the difference between two sets, it is impossible to transform a problem with an unknown reference set into a problem with an unknown compare set and then solve it. However, the results of Experiment 5 do not show conclusively that the reason for the children's failing to solve problems with an unknown reference set is the lack of flexible access to the language describing the difference between sets, because the subjects in this experiment were not tested on solving arithmetic word problems. As my hypothesis was that performing correctly in picture problems with two correct statements concerning the difference between two sets is closely related to solving unknown reference set problems, both tasks needed to be investigated in the same sample at the same time.

Experiment 6: Relation Between Performance in the Picture Task and Unknown Reference Set Problems

Experiment 6 was designed to show that there is a relation between correctly performing the "difference set, both correct" picture tasks and solving unknown reference set problems that is due not only to the difficulty of both tasks but to common conceptual deficits in understanding the symmetry of comparison. Thus, in addition to these tasks, word problems that did not deal with the comparison of sets (but that were similar in difficulty to unknown reference set problems) were presented.

Method

Subjects. Twenty-six first graders (mean age = 7 years, 7 months) and 15 second graders (mean age = 8 years, 4 months) from an after-school center in Munich, Federal Republic of Germany, participated in the study.

Materials and procedure. The subjects were given the picture stories described in Experiment 5. Because I was interested only in "both correct" statement pairs concerning the difference set, we presented the statement pairs Dcm–Dcf and Dcf–Dcm. As a warm-up, the children were given two other statement pairs first. In the same session, the children were to solve two compare problems with an unknown reference set (Problems 3 and 4), two compare problems with an unknown compare set (Problems 1 and 2), and two *change problems with an unknown start set*, as in the following example:

10. In the beginning, Paul had some marbles.
Then Peter gave him 2 other marbles.
Now Paul has 6 marbles.
How many marbles did Paul have in the beginning?

The difficulty of this kind of problem is similar to that of compare problems with an unknown reference set, as the data from Riley and Greeno (1988) and Stern (1992) indicate. In both studies, fewer than 30% of the first graders were able to solve these problems. However, because these problems have nothing to do with the comparison of sets, the relation between performance in the picture story task and performance in the change problems was expected to be lower than the relation between performance in the picture story task and performance on unknown reference set problems. The procedure for problem solving was the same as in Experiment 3. The children were tested individually in the after-school center in a separate room by a male research assistant. Half of the subjects worked on the word problems first and then solved the picture story task, and half of the children performed the tasks in the reverse order. There was a break of about 1 hr between the two tests.

Results and Discussion

To determine whether an arithmetic word problem had been solved, the same criteria as in Experiments 3 and 4 were considered. The results of the picture task were analyzed in the same way as in Experiment 5.

Thirteen of the 15 second graders performed correctly on all tasks of the picture story, and their performance on all arithmetic word problems was close to ceiling: The solution rate for unknown compare set problems was 1.00, for unknown reference set problems it was .80, and for change problems with an unknown start set it was .93. The 2 children who did not perform well on the "both correct" picture task also failed to solve the problems with the unknown reference set.

The second graders' data provide some evidence that performing well on the picture task and solving unknown reference set problems are related. However, because the children also performed well on change problems with an unknown start set, the results do not conclusively show that performing well in the "both correct" picture task is the prerequisite for solving unknown reference set problems.

The first graders' results on the picture test showed that the children either solved both tasks correctly (38% of the children) or solved neither of the tasks. The solution rates for the word problems for the subjects who did and did not pass the picture test are depicted in Table 7.

The correlations between performance on the "difference set, both correct" picture story task and the word problems are .56 for unknown compare set problems ($p < .001$), .62 for unknown reference set problems ($p < .001$), and .16 (*ns*) for change problems with an unknown start set. The test proposed by Olkin (1966) for correlated correlation coefficients showed that the difference between the latter two coefficients was significant ($z = 2.69$, $p < .05$). Therefore, the hypothesis that performance on the picture task is more strongly related to performance on the unknown reference set problems than to performance on other problems of similar difficulty, such as change problems with unknown

Table 7
Mean Solution Rates on Word Problems of First Graders Who Did and Did Not Perform Correctly on the Picture Test in Experiment 6

Picture test performance	Word problems: Compare set	Reference set	Change, unknown start set
Correct ($n = 10$)	1.00	.60	.45
Incorrect ($n = 16$)	.59	.06	.28
Total	.79	.33	.36

start sets, is confirmed. The results of Experiment 6 indicate that compare problems with an unknown reference set are difficult because children do not possess knowledge about the symmetry of comparison.

General Discussion

Why are unknown reference set problems more difficult than unknown compare set problems? The results of Experiments 1–3 suggested that the use of the personal pronoun is not the source of children's difficulties. In Experiment 4 it was shown that most first graders were able to retell and thus understand compare problems with an unknown compare set but were not able to retell compare problems with an unknown reference set. There is no reason to assume that the use of key word strategies is responsible for better performance in solving unknown compare set problems than in solving unknown reference set problems.

All results for first graders indicate that most of these subjects are able to understand the language expressions *more* and *fewer* in a quantitative comparison context. In Experiments 2, 3, and 4, about 50% of the children were able to understand and solve problems with an unknown compare set. The results of the picture story in Experiment 5 further indicate that children could discriminate between true and false statements describing the quantitative difference between two sets even though they did not seem to understand that the quantitative difference between the same sets could be expressed in parallel ways with *both* the terms *more* and *fewer.*

Unknown reference set problems can be solved either by transforming the problem text into part–whole relations, as Riley and Greeno (1988) proposed (mathematical transformation), or by transforming them into unknown compare set problems, as Lewis and Mayer (1987) suggested (linguistic restructuring). In any case, flexibility in the use of language describing quantitative comparison is required. However, the results of Experiment 5 indicate that most of the first graders did not have knowledge about the symmetry of language about quantitative comparison. Rather, for most of the first graders the linguistic representations of "*n* fewer than" and "*n* more than" seemed quite independent of each other, and the activation of one concept seemed to inhibit the activation of the other. The children might have overgeneralized the contrastive meaning of *more* and *fewer* and may have created rules such as "If in one sentence the word *more* is used and in the other sentence the word *fewer* is used, these sentences cannot have the same meaning."

The results of Experiment 6 show that there is a relation between understanding the symmetry of comparison and solving unknown reference set problems that not only is due to the difficulty of both tasks but seems to be caused specifically by common conceptual deficits in understanding the symmetry of comparison.

The awareness of several possibilities for putting the same event into different words is an important criterion for understanding a concept not only in mathematics but also in other domains (Karmiloff-Smith, 1979, 1986). Possibly, not knowing that one can use the expressions "*n* more *x* than *y*" and "*n* fewer *y* than *x*" interchangeably arises from young children's restricted conceptual representations of addition and subtraction. When they enter school, children are able to add and subtract small numbers by using counting procedures, but their knowledge about addition may be restricted to "getting something," and their knowledge about subtraction, to "taking something away." Fuson (1984, 1988), Baroody (1987), and Steffe and Cobb (1988) showed that children interpret subtraction as "take away" and that their preferred strategy is to count backward (e.g., $7 - 5 = ?$ is solved by counting back from 7 five times: 6, 5, 4, 3, 2). Children do not use the "count on" procedure ($7 - 5 = ?$ is solved by counting on from 5 [6, 7] and noting how often one has to count), as has been shown by Carpenter and Moser (1983). This indicates that young children understand both addition and subtraction as independent processes, rather than as reciprocal operations. This deficit in conceptual understanding may be responsible for the children's inflexibility in the use and understanding of the language expressions related to these operations. That is because addition and subtraction are represented as two separate operations that exclude each other, and the language expressions associated with these operations, "*n* fewer than" and "*n* more than," are also separately represented. If children are not aware of the fact that addition and subtraction are different sides of the same coin, then there is no reason for them to see the verbal expressions connected with these mathematical operations—*more* with addition and *fewer* with subtraction—as reciprocal.

In previous articles, (Stern, 1989, 1992), I found that elementary school children showed an amazing inflexibility in mathematizing word problems. Most of them did not know that a word problem could be mathematized in different ways—for example, that Problem 1 can be mathematized by $5 - 2 = 3$, $5 - 3 = 2$, $3 + 2 = 5$, or $2 + 3 = 5$. For compare problems with an unknown compare or reference set, Stern (1989) showed that only a few children used direct mathematization by adding or subtracting the difference set. Even those children who performed very well on other tests of mathematics chose a mathematization that was realized by a fill-in-the-blank equation: large set − small set = difference set. In Stern (1989) I claimed that the children derived this equation from a premathematical match–separate strategy that is used when sets have to be compared (Kintsch & Greeno, 1985). Only when numbers are represented in terms of relations of parts and wholes can the complementary relations of addition and subtraction be understood (Baroody, 1987; Resnick & Greeno, 1990). Having this kind of flexibility in using formal language to describe mathematical facts may also enable children to be flexible in describing quantitative relations with the help of everyday language.

References

Baroody, A. J. (1987). *Children's mathematical thinking.* New York: Teachers College Press.

Briars, D. J., & Larkin, J. H. (1984). An integrated model of skill in solving elementary word problems. *Cognition and Instruction, 1,* 245–296.
Carey, S. (1982). Semantic development: The state of art. In E. Wanner & L. R. Gleitman (Eds.), *Language acquisition: The state of art* (pp. 347–389). Cambridge, England: Cambridge University Press.
Carpenter, T. P., & Moser, J. M. (1983). The acquisition of addition and subtraction concepts. In R. Lesh & M. Landau (Eds.), *Acquisition of mathematics concepts and processes* (pp. 7–44). San Diego, CA: Academic Press.
Clark, E. V. (1973). Non-linguistic strategies and the acquisition of word meanings. *Cognition, 2,* 161–182.
Clark, H. H. (1969). Linguistic processes in deductive reasoning. *Psychological Review, 76,* 387–404.
Cummins, D., Kintsch, W., Reusser, K., & Weimer, R. (1988). The role of understanding in solving word problems. *Cognitive Psychology, 20,* 405–438.
Davis-Dorsey, J., Ross, S. M., & Morrison, G. R. (1991). The role of rewording and context personalization in the solving of mathematical word problems. *Journal of Educational Psychology, 83,* 61–68.
De Corte, E., & Verschaffel, L. (1985). Beginning first graders' initial representation of arithmetic word problems. *Journal of Mathematical Behavior, 4,* 3–21.
De Corte, E., & Verschaffel, L. (1987). 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.
Dellarosa, D. (1986). A computer simulation of children's arithmetic word problem solving. *Behavior Research Methods, Instruments, and Computers, 18,* 147–154.
Donaldson, M., & Balfour, G. (1968). Less is more: A study of language comprehension in children. *British Journal of Psychology, 59,* 461–471.
Fuson, K. C. (1984). More complexities in subtraction. *Journal for Research in Mathematics Education, 15,* 214–225.
Fuson, K. C. (1988). *Children's counting and concepts of number.* New York: Springer.
Greeno, J. G. (1989). Situations, mental models, and generative knowledge. In D. Klahr & K. Kotovsky (Eds.), *Complex information processing* (pp. 285–318). Hillsdale, NJ: Erlbaum.
Hegarty, M., Mayer, R. E., & Green, C. E. (1992). Comprehension of arithmetic word problems: Evidence from students' eye fixations. *Journal of Educational Psychology, 84,* 76–84.
Hudson, T. (1983). Correspondences and numerical differences between disjoint sets. *Child Development, 54,* 84–90.
Huttenlocher, J., & Strauss, S. (1968). Comprehension and a statement's relation to the situation it describes. *Journal of Verbal Learning and Behavior, 7,* 300–304.
Jaspers, M. (1991). *Prototypes of computer assisted instruction for arithmetic word-problem solving.* Unpublished doctoral dissertation, Institute of Special Education, Catholic University of Nijmegen, Nijmegen, The Netherlands.
Johnson-Laird, P. N. (1983). *Mental models.* Cambridge, MA: Harvard University Press.
Karmiloff-Smith, A. (1979). *A functional approach to child language.* Cambridge, England: Cambridge University Press.
Karmiloff-Smith, A. (1986). From meta-processes to conscious access: Evidence from children's metalinguistic and repair data. *Cognition, 23,* 95–147.
Kintsch, W. (1988). The role of knowledge in discourse comprehension: A construction–integration model. *Psychological Review, 95,* 163–182.
Kintsch, W., & Greeno, J. G. (1985). Understanding and solving word arithmetic problems. *Psychological Review, 92,* 109–129.
Klatzky, R. L., Clark, E. V., & Macken, M. (1973). Asymmetries in the acquisition of polar adjectives: Linguistic or conceptual? *Journal of Experimental Child Psychology, 16,* 32–46.
Lewis, A. B., & Mayer, R. E. (1987). Students' miscomprehension of relational statements in arithmetic word problems. *Journal of Educational Psychology, 79,* 363–371.
Morales, R. V., Shute, V. J., & Pellegrino, J. W. (1985). Developmental differences in understanding and solving simple mathematics word problems. *Cognition and Instruction, 2,* 41–57.
Nesher, P., & Teubal, E. (1975). Verbal cues as an interfering factor in verbal problem solving. *Educational Studies in Mathematics, 6,* 41–51.
Oakhill, J., & Yuill, N. (1986). Pronoun resolution in skilled and less-skilled comprehenders: Effects of memory load and inferential complexity. *Language and Speech, 29,* 25–37.
Olkin, I. (1966). Correlations revisited. In J. C. Stanley (Ed.), *Improving experimental design and statistical analysis* (pp. 102–156). Chicago: Rand McNally.
Resnick, L. B. (1989). Developing mathematical knowledge. *American Psychologist, 44,* 162–169.
Resnick, L. B., & Greeno, J. G. (1990). *Conceptual growth of number and quantity.* Unpublished manuscript, University of Pittsburgh.
Reusser, K. (1989). *Textual and situational factors in solving mathematical word problems* (Research Rep. No. 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. Bennett, & H. F. Friedrich (Eds.), *Learning and instruction: Vol. 2.2. Analysis of complex skills and complex knowledge domains* (pp. 477–498). Elmsford, NY: 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. H. (1983). Development of children's problem-solving ability in arithmetic. In H. P. Ginsburg (Ed.), *The development of mathematical thinking* (pp. 153–196). San Diego, CA: Academic Press.
Schoenfeld, A. H. (1982). Some thoughts on problem-solving research and mathematics education. In F. K. Lester, Jr., & J. Garofalo (Eds.), *Mathematical problem solving: Issues and research* (pp. 27–37). Philadelphia: Franklin Institute Press.
Siegler, R. S. (1981). Developmental sequences within and between concepts. *Monographs of the Society for Research in Child Development, 46*(2, Serial No. 189).
Steffe, L. P., & Cobb, P. (1988). *Construction of arithmetical meanings and strategies.* New York: Springer.
Stern, E. (1989, September) *The role of arithmetic in solving word problems.* Paper presented at the 3rd Conference of the European Association of Research in Learning and Instruction, Madrid, Spain (Unpublished Paper No. 20, Max-Planck-Institute for Psychological Research, Munich, Federal Republic of Germany).
Stern, E. (1992). Warum werden Kapitänsaufgaben "gelöst"? Das Verstehen von Textaufgaben aus psychologischer Sicht [Why do children solve nonsense problems? Understanding and solving arithmetic word problems from a psychological point of view]. *Der Mathematikunterricht, 4,* 7–29.
Stern, E., & Lehrndorfer, A. (1992). The role of situational context in solving word problems. *Cognitive Development, 2,* 259–268.

Stern, E., & Schneider, W. (1989). Development of children's understanding of number between the age of four and six. In F. E. Weinert & W. Schneider (Eds.), *The Munich Longitudinal Study on the Genesis of Individual Competencies (LOGIC). Report No. 6: Psychological development in the preschool years: Longitudinal results of wave one to three* (pp. 14–19). Munich, Federal Republic of Germany: Max-Planck-Institute for Psychological Research.

van Dijk, T. A., & Kintsch, W. (1983). *Strategies of discourse comprehension.* San Diego, CA: Academic Press.

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

Wilcox, S., & Palermo, D. (1977). *In, on, under; more, less; some artifacts revealed.* Paper presented at the meeting of the Eastern Psychological Association, Boston.

Wykes, T. (1981). Inference and children's comprehension of pronouns. *Journal of Experimental Child Psychology, 32,* 264–278.

Received October 25, 1991
Revision received July 3, 1992
Accepted July 3, 1992 ■