COGNITION AND INSTRUCTION, 1991, *8*(3), 261–289

Children's Interpretations of Arithmetic Word Problems

Denise Dellarosa Cummins
Yale University

Two experiments were conducted to investigate children's interpretations of standard arithmetic word problems and the factors that influence their interpretations. In Experiment 1, children were required to solve a series of problems and then to draw and select pictures that represented the problems' structures. Solution performance was found to vary systematically with the nature of the representations drawn and chosen. The crucial determinant of solution success was the interpretation a child assigned to certain phrases used in the problems. In Experiment 2, solution and drawing accuracy were found to be significantly improved by rewording problems to avoid ambiguous linguistic forms. Together, these results imply that (a) word-problem solution errors are caused by misinterpretations of certain verbal expressions commonly used in problem texts, and (b) these misinterpretations are the result of missing or inadequate mappings of these verbal expressions to part-whole knowledge.

An important component of mathematics training is solving word problems. Real-world problems that require mathematics for their solution typically do not come to us as equations ready to be solved but rather as verbal or pictorial representations that must be interpreted symbolically, manipulated, and solved. It is for this reason that word problems are introduced in the earliest stages of mathematics instruction.

Interestingly, children's performance on arithmetic word problems varies as a function of the format of the problem and the age of the child (Carpenter, Hiebert, & Moser, 1981; Cummins, Kintsch, Reusser, & Weimer, 1988; De Corte & Verschaffel, 1985, 1987; De Corte, Verschaffel, & De Win, 1985; Riley & Greeno, 1988; Riley, Greeno, & Heller, 1983). In attempting to understand this variability, researchers have typically focused on two crucial questions: (a) What knowledge is necessary to solve various

Requests for reprints should be sent to Denise Dellarosa Cummins, Department of Psychology, University of Arizona, Tucson, AZ 85721.

TABLE 1
Examples of Problems Used

Combine 5[a]
Mary and John have 5 marbles altogether.
Mary has 3 marbles.
How many does John have?
Compare 4[b]
Mary has 7 marbles.
John has 4 marbles less than Mary.
How many marbles does John have?
Compare 6[b]
Mary has 8 marbles.
She has 2 marbles less than John.
How many marbles does John have?
Change 6[b]
Mary had some marbles.
Then she gave 2 marbles to John.
Now Mary has 6 marbles.
How many marbles did Mary have in the beginning?

Note. Based on Riley, Greeno, and Heller's (1983) classification.
[a]Used in both experiments. [b]Used in Experiment 1 only.

types of word problems? and (b) How is this knowledge acquired? To answer these questions, some investigators have suggested treating word problems as valid discourse entities and have offered analyses of the types of knowledge that must be accessed and used in order to interpret and comprehend problem texts. One of the best known treatments is that of Riley and her colleagues (Riley & Greeno, 1988; Riley et al., 1983). They offer a competence model of simple arithmetic word-problem solving in which solution success depends crucially on understanding the *semantic relations* described by the text: "By *semantic relations* we refer to the conceptual knowledge about increases, decreases, combinations, and comparisons involving sets of objects" (Riley et al., 1983, p. 159). Simple arithmetic word problems can be classified in terms of the types of semantic relations they describe. Riley et al. (1983) argued that many standard problems can be classified into just three semantic types: *combine, change,* and *compare* problems. Combine and compare problems describe static relations between quantities, whereas change problems describe actions that cause increases or decreases in some quantity. Examples of these problems are provided in Table 1. In the combine problem, two quantities must be considered in combination to produce the answer (e.g., How many altogether?). In the compare problems, two quantities must be compared and their difference quantified. In the change problem, an exchange occurs that changes the size of a given set of objects.

According to Riley et al. (1983), a competent problem solver possesses schemata that represent these semantic relations and connect them to solution sequences. Within this framework, problem-solving skill improves in direct relation to increases in the child's representational capacity:

> We hypothesize that acquisition of <problem solving> skill is primarily an improvement in children's ability to understand problems—that is, in their ability to represent the relationships among quantities described in the problem situation in a way that relates to available solution procedures. (p. 173)

or more particularly

> In our analysis the main locus of children's improvement in problem-solving skill is in the acquisition of <problem type> schemata for understanding the problem in a way that relates it to already available action schemata. (p. 171)

This analysis of semantic structures has proved useful and has been readily adopted by a variety of researchers in the field (e.g., Cummins et al., 1988; De Corte et al., 1985; Kintsch & Greeno, 1985). In answer to the second question (How is this knowledge acquired?), Riley and her colleagues argue that semantic schema acquisition itself depends on *acquisition of knowledge concerning logical set relations, particularly part–whole relations.* Riley et al. described three models of problem solving that represent three levels of competence in solving change problems. In describing the differences between their low competence model on the one hand, and their middle and high competence models on the other, Riley et al. (1983) stated:

> Models (2) and (3) also have . . . a richer understanding of certain relations between numbers; for example, model (3) has an understanding of part–whole relations. (p. 176)

> In fact, it is this model's lack of understanding of part–whole relations that accounts for its failure on combine 2 problems. (p. 184)

> Furthermore, there is evidence suggesting that *once children understand part–whole relations* [italics added], they can use this knowledge to understand all change problems. (p. 181)

In other words, the crucial process that drives development of problem-solving schemata and, hence, skill is the *acquisition of knowledge concerning part–whole relations.* A fully competent problem solver understands and represents word problems as relations among parts and wholes. Semantic schemata identify which problem quantities are parts and which are wholes. They have been described as part–whole schemata that are contextualized to map smoothly onto standard arithmetic story-problem structures (Cummins et al., 1988). More recently, Riley and Greeno (1988) extended this account of skill and schema development to encompass combine and compare problems as well. They suggested, for example, that understanding comparative phrase forms such as "*X* has *n* more marbles than *Y*" requires understanding that numbers can be values for operators and not just set cardinalities. *Hence, a child's logical and mathematical knowledge is presumed to drive understanding of the semantics of the problem text.*

This *logico-mathematical* explanation of children's skill development traces its roots back to Inhelder and Piaget's (1964) observations of children's reasoning about class inclusion terms (e.g., Are there more daisies or more flowers?). Inhelder and Piaget concluded that young children fail class inclusion problems because they do not understand part–whole relations. Proponents of the logico-mathematical development view use similar arguments to explain children's poor performance on certain arithmetic word problems. According to Riley's analysis, poor performance on certain word problems reflects a lack of sufficient knowledge concerning part–whole set relations. Acquisition of such knowledge impacts on problem-solving skill indirectly by improving the solver's representational schemata. Other proponents of this view suggest that acquisition of such knowledge impacts more directly on problem-solving solution strategies. For example, Nesher (1982) attributed varying levels of performance on word problems to the development of knowledge concerning sets, logical operations, and mathematical operations. Similarly, in the model of skill development proposed by Briars and Larkin (1984), acquisition of logical knowledge, such as the reversibility of certain operations in time, subset equivalence, and set cardinality, directly determines the strategies chosen by the problem solver. Describing the least skilled version of their model, CHIPS, the authors stated:

> However, single-role CHIPS has no knowledge of sets—just arrays of counters. The Riley model, even Model 1, represents problems by the quantities in the problem, that is by the cardinal numbers of the sets involved. *Thus we believe that CHIPS stays somewhat closer to the primitive abilities of very young children* [italics added]. (p. 288)

The most skilled model depends on critical conceptual knowledge in order to perform maximally:

> CHIPS using subset equivalence and time-reversal schemas performs equivalently to Riley's Model 3, using its part–whole schema (Riley, Greeno, & Heller, 1983). This CHIPS, like Riley's model, represents problems with sets and their cardinals, rather than by individual items. The major difference between the models at this level is CHIPS' ability to use the time-reversal and subset equivalence schemas as an alternative solution method. (Briars & Larkin, 1984, p. 289)

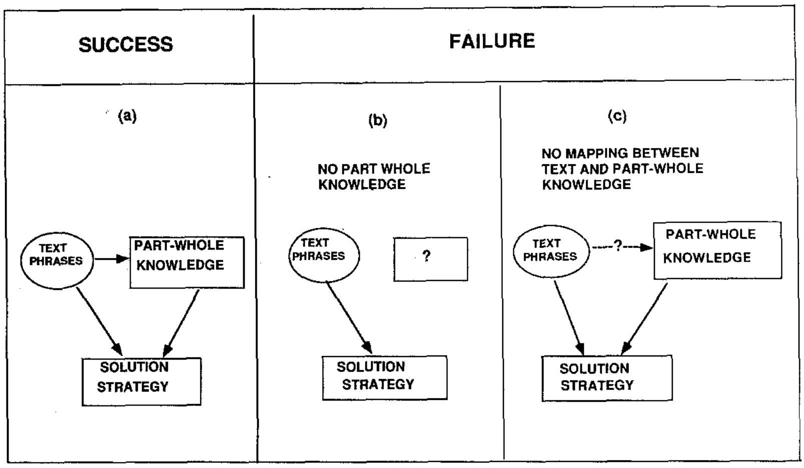


FIGURE 1 Explanatory frameworks for successful and for poor problem-solving performance. (a) In successful problem solving, words and phrases from the problem text either directly cue solution procedures or are mapped onto part-whole set structures, which in turn cue solution procedures. (b) According to the logico-mathematical viewpoint, solution failures result from a lack of sufficient part-whole knowledge. If the text does not contain phrases that directly cue solution procedures, the problem cannot be solved, because the alternative route cannot be taken. (c) According to the linguistic development view, solution failures result from missing or inadequate mappings of verbal expressions to part-whole structures, leading to a failure to access part-whole knowledge. Rewording enhances performance by increasing the probability that part-whole knowledge will be accessed successfully.

One way of depicting the logico-mathematical view of skill development is presented in Figure 1, Panels a and b. Panel a depicts the competent problem solver's solution strategies. A word problem can be solved by mapping simple key words onto procedures (e.g., *altogether* onto *count*) or by mapping the text propositions onto part-whole knowledge, which in turn triggers a solution procedure. The latter mappings constitute a semantic schema in which the set relations described by the problem text are coherently specified. Panel b depicts a problem solver of lesser skill. Here, only the direct mappings of key words onto procedures are available; the schematic mappings of text propositions onto part-whole knowledge cannot be accomplished (or not completely) due to the lack of sufficient part-whole knowledge. The transition between the two levels of competence depends crucially on the development of knowledge concerning part-whole relations.

A major problem with the logico-mathematical development view is that empirical evidence suggests children often know more about logical set relations than the view supposes. For example, there is a great deal of evidence that middle-class children possess a fairly good understanding of cardinality by age 4, contradicting one of the assumptions on which Briars and Larkin based their model of school-age problem-solving skill (see Fuson & Hall, 1983, for a review). Other evidence suggests that preschoolers possess at least a tacit understanding of part–whole relations and subset equivalence (Hudson, 1983; Markman, 1973; Markman & Siebert, 1976; Smith & Kemler, 1978; Trabasso et al., 1978). Interestingly, the key to observing such tacit knowledge in these studies depends on how the problem task is worded. For example, even preschoolers can solve class inclusion problems (which require knowledge of part–whole relations) when terms that emphasize the aggregate nature of the set are used (Markman, 1973; Markman & Siebert, 1976). Similarly, 4-year-olds have been found to use solution strategies that evidence knowledge of part–whole relations and subset equivalence on reworded comparison problems. For example, Hudson (1983) presented 4-year-olds with comparative word problems accompanied by pictures. Few of the children were capable of solving such problems when they were worded in standard form, such as "There are 5 birds and 3 worms. How many more birds are there than worms?" Solution performance increased dramatically, however, when the last line of the problem was reworded as "How many birds won't get a worm?" More important, the solution strategies children employed evidenced a tacit understanding of logical set relations. The most common strategy the children used involved counting the number of worms (on the accompanying picture), counting out a subset of birds equal to the cardinality of worms, and returning the cardinality of the remaining subset of birds as the answer. This strategy implies a tacit understanding of one-to-one correspondence and subset equivalence of sets with identical cardinalities, as well as the part–whole structure of the sets in question. Similar effects of rewordings on solution performance of school-age children were reported by De Corte et al. (1985).

Interestingly, proponents of the logico-mathematical view have cited results like these as supporting their view. For example, when discussing such results, Riley et al. (1983) stated,

> The point to be made here is that once again we cannot attribute children's problem-solving difficulties to a deficiency involving problem-solving *actions* [italics added]. Instead, we hypothesize that acquisition of skill is primarily an improvement in children's ability to understand problems—that is, in their ability to represent the relationships among quantities described in problem situations in a way that relates to available solution procedures. (p. 173)

In other words, Riley et al. (1983) cited rewording results as evidence that children possess adequate solution procedures for solving such problems but fail to enact them because they fail to understand the problem properly. Again, what drives this understanding, however, is development of conceptual knowledge concerning part–whole relations. Children do not access the correct solution procedure because they fail to understand the problem in terms of part–whole relations, and they fail to understand the problem in those terms because they suffer from a lack of knowledge concerning such relations. The models just described carefully track the development or acquisition of such knowledge.

The difficulty with this view is that it is not clear how a simple rewording can allow a child without sufficient part–whole knowledge to solve such problems. In other words, why should a simple rewording matter? One possibility is that the rewording simplifies the problem structure so that part–whole knowledge is not required to solve it; that is, the rewording provides key words or the like that immediately trigger a problem-solving action. As Hudson (1983) was at pains to point out, however, this is not the case. Even 4-year-olds exhibited a sophisticated grasp of part–whole relations in their problem-solving procedures. The data instead seem to indicate that the knowledge is there, but it simply is not accessed when problems are worded in certain ways.

To accommodate such evidence, some researchers have proposed that a major component of problem-solving skill development is the acquisition of information concerning the interpretation and use of language in word-problem contexts. De Corte and Verschaffel (1985), for example, suggested that skill development depends on the acquisition of Riley-like semantic schemata and on a more general "word-problem schema" that indicates the structure, role, and intent of word problems in general. The main purpose of word-problem schemata is to encode implicit rules, suppositions, and agreements concerning typical word problems that will enable a solver to interpret ambiguities correctly and to compensate for insufficiencies in the verbal text. In a sense, the word-problem schema is a sort of culturally determined "shared knowledge" (in a Gricean sense) that allows a solver to assign correct interpretations to impoverished problem texts. It embodies the pragmatics of the word-problem domain, whereas the schemata proposed by Riley embody the semantics of problem classes in the domain.

Like De Corte and Verschaffel, Cummins et al. (1988) proposed that language difficulties underlie poor solution performance. They proposed, however, a more localist interpretation of such difficulties. They suggested that certain words and phrases are ambiguous to a child (or other English language novice) and that use of such terms in word problems leads to incoherent representations. The first of these terms is *comparatives*. Cummins et al. argued that children's poor performance on class inclusion problems and word problems containing such terms indicates that children often have difficulty interpreting them. This view is supported by the fact that children often transform them into simple possession terms when retelling word problems (Cummins et al., 1988, Experiment 1), skip over them when reading phrases containing them (De Corte & Verschaffel, 1986), and perform better when problems containing them are reworded to exclude them (De Corte et al., 1985; Hudson, 1983). Like Mayer (1982), they suggested that faulty problem representations result from children's interpreting comparative terms as statements of simple possession or simple assignment. Thus, "Mary has 5 more marbles than John" is interpreted as "Mary has 5 marbles." They reported the results of a computer simulation of children's problem solving in which the simulation's solution and error protocols best matched children's when its lexicon was altered to interpret comparatives in just this way.

Another term that Cummins et al. (1988) suggested proves problematic for children is *altogether* when it is used with conjunctives to indicate joint ownership, as in "Mary and John have 5 altogether." Like De Corte and Verschaffel (1985), Cummins et al. suggested that children interpret the conjunction in this context to mean "each," as evidenced in retelling protocols such as the following (I = interviewer, C = child):

I: Pete has 3 apples; Ann also has some apples; Pete and Ann have 9 apples altogether; how many does Ann have?
C: 9.
I: Why?
C: Because you just said it.
I: Can you retell the story?
C: Pete had 3 apples; Ann had also some apples; Ann had 9 apples; Pete also has 9 apples.

(De Corte & Verschaffel, 1985, p. 19)

Using the computer simulation once again, Cummins et al. found that the best match of simulation to observed performance occurred when the simulation's lexicon was altered to allow interpretation of this term as *each*. Finally, Cummins et al. suggested that the term *some*, as in "John had some marbles," is often improperly interpreted by children. As evidence, they showed that the best match of simulation to observed performance occurred when the simulation's lexicon was altered to prevent interpretation of this term as a *quantity* term.

To summarize, the linguistic development view suggests that a major source of difficulty children encounter when solving word problems is properly interpreting certain words and phrases in terms of sets and logical set relations. Although there is some disagreement about whether these mappings constitute pragmatic knowledge or more local (lexical) knowledge, proponents of this view generally agree that *children possess at least tacit understanding of part–whole relations, and what they come to learn through instruction or further familiarization with language is how certain verbal formats map onto those relations.* This view is depicted in Figure 1, Panel c. Here, the solver's knowledge base contains knowledge concerning part–whole relations and knowledge concerning the operational meanings of certain words used in arithmetic word problems (e.g., *less* means "decrement," *more* means "increment"). Missing, however, are the interpretations of certain phrases in terms of part–whole relations. For example, a child may possess, at least tacitly, the concept of part–whole relations yet still be unclear about how the rather baroque phrase, "How many more *X*s are there than *Y*s," maps onto parts and wholes. Or a child may not understand that the term *altogether,* when used in a conjunction phrase such as "Mary and John have 5 marbles altogether," denotes the whole. The development of problem-solving skill from this view depends crucially on acquiring these mappings from words and phrases onto part–whole knowledge, because the mappings indicate how such words and phrases are to be interpreted in the context of arithmetic word problems.

The issue of how children interpret and represent word problems is of particular importance to educators, because remediation depends heavily on identifying where a child's misunderstanding lies and what the nature of the misunderstanding is. The difficulty with the work described so far is that it is based on inferences about children's interpretations based on indirect evidence (e.g., retelling data, quantitative solution performance, or simulation work). A major problem with retelling data is that one is never sure whether incorrect retellings reflect memory storage failures, retrieval failures, or true misunderstandings. The purpose of the work presented here was to gather more direct information about children's interpretations and, in so doing, to test further the two conflicting explanations of children's solution errors. Evidence concerning children's interpretations was gathered in Experiment 1 by requiring children to draw and choose pictures illustrating situations described in word problems that used comparatives and terms such as *altogether* and *some.* These drawings and picture choices were then compared with the children's solution strategies. If, as hypothesized, children do misinterpret statements including these terms, these misinterpretations should be evidenced by children's drawings and picture choices. Moreover, there should be systematic relationships between their misinterpretations and their solution performance.

The purpose of Experiment 2 was to rule out the possibility that such misinterpretations occur because the children making them do not possess sufficient conceptual knowledge of part–whole relations to allow them to interpret such forms properly. This was done by choosing children who consistently evidenced interpretation and solution errors on such problems and observing their performance on reworded versions of the problems. If these children produced incorrect depictions and solutions of standard part–whole problems and correct depictions of reworded part–whole problems, their performance differences could not be due to deficient part–whole conceptual knowledge.

EXPERIMENT 1

The purpose of this experiment was to gather direct evidence of children's interpretations of four standard word problems. Examples of these problems are presented in Table 1. First graders were required to solve a series of these problems. Twenty-four hours later, their interpretations of the problem situations were probed in the following ways. First, they were required to draw a representation of important aspects of each type of problem. Second, they were required to choose pictures that best represented the situations described in the story. These two measures, therefore, represented a child's ability to generate and to recognize a representation of the problem situation, respectively. It was predicted that a systematic relationship would be observed between children's interpretations as depicted in drawings and picture choices, and strategies they adopted on the solution task. Specifically, it was expected that children who solved combine-5 problems incorrectly would be those who misinterpreted the first line of the problem to mean that Mary and John *each* have the stated number of marbles. Children who solved compare-4 and compare-6 problems incorrectly were expected to misinterpret the comparative verbal form as a simple statement of possession, producing a "stimulus-matching" strategy wherein one number is associated with Mary and the other with John. Finally, children who solved change-6 problems incorrectly were expected to misinterpret or ignore the term *some* in the first line of the problem.

Method

Subjects

Twenty-four first-grade children (10 girls, 14 boys) from the New Haven Public Schools served as subjects in the experiment. The children were middle class and racially mixed. The children were tested late in the school year (March). Written parental consent was obtained for each child's participation, and the schools were paid $5.00 for each child who participated.

Materials

Examples of the four types of problems used in the study are presented in Table 1. Two of the problem types required addition (compare 6 and change 6), and the other two required subtraction (combine 5 and combine 4). The children were required to solve three instances of each type of problem plus one practice problem for a total of 13 problems. The number triples embedded in these problems were chosen so that correct answers were less than 10 and were not the same as numbers in the problem. Also, to ensure that solution performance on the problem types was not an artifact of differences in difficulty of the numbers embedded in them, two sets of number pairs were constructed and assigned to the problems. Number set was treated as a between-subjects variable. This meant that a given problem (e.g., the first instance of combine 5) was tested using one number pair for half the children and another number pair for the remaining half of the children. Two presentation orders were also used and were counterbalanced with problem sets across subjects.

For the drawing task, children were given a template (illustrated in the Appendix) in which to draw marbles. For the picture selection task, materials were constructed in the following way: One target picture and two foils were constructed for each line of each problem. The pictures used for one of the number sets are included in the Appendix, arranged so that the target is always the first picture in the set. Presentation order was randomized for the actual materials used in the experiment. The target for lines that simply stated ownership (e.g., Mary has five marbles) showed a girl kneeling in front of an ellipse that contained five marbles. The foils included a boy kneeling in front of five marbles and a girl kneeling in front of an arbitrary number of marbles (e.g., seven). Crucial lines for the problems were depicted as described in the following four paragraphs.

Combine 5. The target for "Mary and John have 5 marbles altogether" depicted a boy and girl kneeling in front of five marbles. One foil included a boy and girl kneeling in front of 10 marbles, 5 in front of the girl and 5 in front of the boy. This was a depiction of the hypothesized incorrect "each" interpretation. The second foil depicted a boy and girl kneeling in front of an arbitrary number of marbles (e.g., seven marbles). For some of these foils, the arbitrary number was represented as a single set (e.g., seven marbles); for others, it was represented as two sets (e.g., seven in front of Mary and seven in front of John). Across subjects, the number of these alternative types of second foils was equated.

Compare 4. The target and foils for "John has 2 marbles less than Mary" all depicted a girl kneeling in front of the number stated in the first line of the problem. Next to her was a boy kneeling in front of a set of marbles. For the target, the number of marbles in front of the boy was two less than that in front of the girl. For the first foil, the number of marbles in front of the boy was the same as that stated in the second line of the problem (e.g., two); this was a depiction of the hypothesized "stimulus-matching" misinterpretation in which the comparative form is simply treated as a statement of possession. For the second foil, the number of marbles for the boy was the sum of the two numbers in the problem.

Compare 6. The target and foils for "She has 2 marbles less than John" all depicted a girl kneeling in front of the number stated in the first line of the problem. Next to her was a boy kneeling in front of a set of marbles. For the target the number of marbles in front of the boy was two more than that in front of the girl. For the first foil, the number of marbles in front of the boy was the same as that stated in the second line of the problem (e.g., two); this was a depiction of the hypothesized "stimulus-matching" misinterpretation. For the second foil, the number of marbles for the boy was the difference of the two numbers in the problem.

Change 6. The crucial lines were the first two. The target for the first line, "Mary had some marbles," depicted a girl kneeling in front of a bag of marbles. (The children were informed that the bag had an unknown number of marbles in it.) The first foil depicted a girl kneeling in front of an arbitrary number of marbles, and the second depicted a boy kneeling in front of an arbitrary number of marbles. The target for the second line, "Then she gave John two marbles," depicted a girl holding an open bag of marbles and handing two marbles to the boy. The first foil was the same except that the girl was not holding an open bag, representing a failure to interpret *some* as an unknown quantity that must be carried forward in the problem. The second foil depicted the boy handing two marbles to the girl. The presence and absence of the bag in the pictures were pointed out to the children without further comment.

Practice problem. To ensure that the children understood the task, a set of targets and foils was constructed for an easy type of problem to use as a practice problem. ("Mary has 5. John has 3. How many altogether?") The target for the last line showed, for example, eight marbles, and the foils showed arbitrary numbers.

Procedure

Children were tested in pairs. They were seated on either side of a divider that prevented them from seeing or conversing with each other but that allowed the experimenter to see and converse with them individually, if needed. On the first day of testing, the problems were read to them and they were required to solve them. Each problem was read twice. The children were given paper and pencil to use in solving the problems. (Most forewent using the paper and pencil in favor of counting on their fingers.) Child pairs were randomly assigned to number set and presentation order conditions. On the second day, the children were given a booklet containing the drawing templates and the picture selections. They performed the drawing task first, followed by the picture selection task. The procedures for the two tasks were as follows:

Drawing task: Four problems were used for the drawing task: a combine-5 problem, a compare-4 problem, a compare-6 problem, and a change-6 problem. The problems were read aloud to the child, and drawings were required after the second reading. The numbers in these problems were not the same as any used the previous day. For the combine-5 problem, the children were asked to draw the number of marbles that went with the first line of the problem. For the compare problems, they were asked to draw in the number of marbles that Mary had in front of the girl and the number that John had in front of the boy on the template. For the change problem, they were asked to draw what they thought was the best way to show *some* marbles, as in "Mary had some marbles."

Picture selection task: For this task, the problems the children solved the day before were again read to them. The booklet containing the targets and foils for the problems was placed in front of them. After each line, they were required to choose the picture from among the three on the page in front of them that "went with the line best." When they found the line that "went best," they were asked to put a mark on the line next to that picture. After making their selection, they were instructed to turn the page, and the next line was read to them. When the final line of the problem was reached, the entire problem was read to them again. Their picture choice for the final line of the problem represented their solution to the problem when it was presented in picture format. The session always began with the practice problem to ensure that children understood the task.

Results and Discussion

Drawing and Solution Accuracy

Subjects were divided into two groups based on drawing accuracy for each type of problem, except for change 6. (All children simply drew an arbitrary number of marbles to illustrate *some* marbles, so no distinctions could be made among subjects in terms of drawing accuracy for this problem.) Mean solution accuracy for the groups on the remaining problems is presented in Table 2.

Children split nearly evenly on drawing accuracy, with 13 of them drawing correct interpretations for "Mary and John have 5 altogether" and 11 drawing erroneous interpretations. Sixty-four percent of the incorrect drawings were the predicted "each" interpretation in which Mary and John were each assigned five marbles. The remaining 36% of the incorrect drawings were unclassifiable (using a strict criterion) either because (a) they depicted

TABLE 2
Mean Proportion Solution Accuracy as a Function of Drawing Accuracy in Experiment 1

	Drawing Accuracy	
Problem Type	*Incorrect*	*Correct*
Combine 5	.17[a] (.34)	.57[b] (.39)
Compare 4	.33[c] (.39)	.72[d] (.36)
Compare 6	.33[c] (.33)	.62[d] (.36)

Note. Numbers in parentheses are standard deviations. Means are based on three observations from each subject.

[a]$n = 14$. [b]$n = 10$. [c]$n = 11$. [d]$n = 13$.

separate sets of marbles in front of the boy and girl, but the sets contained numbers other than the number stated in the sentence (e.g., 5 and 6, or 4 and 5), or (b) the child had drawn a continuous series of 10 marbles on the inside perimeter of the ellipse beginning at John and ending at Mary. The number of "each" interpretations may be underestimated as a result, because it could be argued that each of the unclassifiable cases suggests this interpretation. In the former case, the child may have intended to draw five marbles in front of both Mary and John but lost count of how many marbles had been drawn. In the latter case, 10 marbles clearly had been drawn, but the division between Mary and John was not made explicit. If these were taken as representations of the "each" interpretation, all misinterpretations would be of this type.

For the compare problems, 11 children drew correct interpretations, and 13 drew incorrect ones. The percentages of incorrect drawings that were stimulus matching (i.e., assigning one number to Mary and the other to John) were 82% for compare 4 and 72% for compare 6. Notice that for compare 6, this suggests that children ignored the pronoun *she* in the second line of the problem and simply assigned one number to the boy and one to the girl. It should be noted that another frequent error children made on these problems was to assign no marbles to John (17%), regardless of whether Mary was assigned the first number mentioned, the second, or (in some cases) both. This error can be readily understood if both the first and second statements in the problem are taken as assignment statements pertaining to Mary. The propensity to divide the numbers between the two characters is consistent with De Corte and Verschaffel's (1985) "word problem schema" view, in that children may assume another character would not be introduced in the story if he or she were not also to be assigned some ownership or transactional role.

As predicted, solution accuracy varied systematically with drawing accuracy, with children who drew correct interpretations outperforming those who drew incorrect ones, $t(22) = 2.68$, $p < .025$; $t(22) = 2.51$, $p < .025$; $t(22) = 1.99$, $p < .06$, for combine-5, compare-4, and compare-6 problems, respectively (see Table 2 for means).

The vast majority of errors were wrong operation errors (adding when subtraction was required by the given problem and vice versa) or given number errors (giving back one of the numbers from the problem as the answer). The majority of these errors were committed by children who drew incorrect interpretations. For combine 5, the mean proportion of given number and wrong operation errors for children who misinterpreted *altogether* as "each" was .42 and .44, respectively; this contrasted sharply with .07 and .27, respectively, for children who evidenced correct interpretations. One may well ask why a child who misinterpreted the first line of the problem to mean that John and Mary had a certain number "each" would make a wrong operation error. The answer may be found in the difference in commission rates for the two types of errors even when the problem was interpreted correctly. Wrong operation errors (i.e., adding the numbers) seems to be the preferred strategy for solving these types of problems. Given uncertainty as to the proper interpretation of the problem, it seems reasonable to assume that children would resort to their preferred default strategy. (Such uncertainty would be introduced if *altogether* were interpreted as "each," because the resulting problem representation is incoherent, and the required solution is unclear.) From this perspective, the difference in proportion of given number errors (.42 vs. .07) is striking. A given number error is not the preferred strategy for combine problems, but it is used more frequently when the problem interpretation clearly dictates it.

Similar results were observed for the compare problems. Here, mean proportions given number and wrong operation errors were .31 and .14, respectively, for children who committed a stimulus-matching interpretation error; again, this contrasted sharply with .04 and .07, respectively, for those who evidenced correct interpretations. Notice that no real preference among error types (defaults) for this type of problem is apparent in the correct interpretation case.

Picture Selection and Solution Accuracy

Mean proportion picture selection accuracy for each type of problem is presented in Table 3, along with mean solution accuracy for the word problems from the previous day. It was predicted that performance on the crucial lines would correlate significantly with solution accuracy on the word problems. To test this prediction, four regression analyses were done, one for each problem type, in which picture choice accuracy was regressed onto word-problem solution accuracy. For combine 5, compare 4, and compare 6, the selection on the first two pictures was used as predictor variables; for

TABLE 3
Mean Proportion Picture Selection Accuracy and Solution Accuracy From Experiment 1

		Picture Selection			
Problem Type	*Solution*	*Picture 1*	*Picture 2*	*Picture 3*	*Picture 4*
Combine 5	.33	.67[a]	.21	.58	—
Compare 4	.54	.99	.50[a]	.64	—
Compare 6	.48	1.00	.68[a]	.72	—
Change 6	.64	.43[a]	.56[a]	1.00	.53

Note. Means are based on three observations from each of 24 subjects. *Solution* is the mean proportion solution accuracy on word problems from the first day of testing. The remaining columns are mean proportion accuracy on the picture selection task. The pictures correspond to each line in the problems. Combine and compare problems have three lines each; change 6 has four.

[a]Crucial pictures for each type of problem.

change 6, selection accuracy on the first three was used. A forward selection procedure was used. As in the previous analyses of drawing accuracy, the results of the regressions support the predictions for three of the four problem types. Solution accuracy on combine-5 problems was found to vary with the ability to select an accurate depiction of the first line of the problem, accounting for 22% of the variance in solution accuracy, $F(1, 22) = 6.26$, $MS_e = .13$, $p < .01$. For compare 4 and compare 6, solution accuracy was accounted for by selection performance on the second line, $Fs(1, 22) = 16.64$ and 4.08, $MS_e = .10$ and $.10$, $ps < .01$ and $.05$, respectively. These regressions accounted for 43% and 16% of the variance, respectively. This is consistent with the results on drawing accuracy, in that the ability to solve these problems varied systematically with the ability to recognize the correct interpretation of the crucial lines of the problems. Also consistent with the results on drawing accuracy is the failure to find systematic variation for change 6. Solution accuracy on change 6 did not correlate with picture selection performance on any of the first three lines. This negative result may indicate either that children simply found the picture format for this task confusing or that they did not yet understand the concept of a variable.

Also analyzed was the relationship between selection performance on the last picture of the problems and selection accuracy on the crucial pictures. (Recall that choosing a picture that best represented the final line of the problem essentially meant solving the problem when it was presented in pictorial form.) To do this, proportion picture selection accuracy on the final lines of the problems was regressed onto proportion picture selection accuracy for the preceding lines of the problems. (For the combine and compare problems, this meant that selection accuracy for the third line was regressed onto accuracy for the first two lines; for the change problems, this meant that performance on the fourth line was regressed onto the first three lines.) The results mirrored those mentioned before. For combine 5, the ability to choose a correct representation of the final line of the problem correlated with the ability to choose a correct representation of the first line of the problem, $F(1, 22) = 10.46$, $MS_e = .12$, $p < .01$, accounting for 32% of the variance. For compare 4 and compare 6, performance on the last line correlated with performance on the crucial second line of the problem, $Fs(1, 22) = 35.80$ and 4.56, $MS_e = .05$ and $.10$, $ps < .001$ and $.05$, respectively. The proportions of variances accounted for were .62 and .17, respectively. Once again, performance on change 6 was not accounted for by performance on any of the previous lines.

Another question addressed was whether children's ability to solve these problems was enhanced by pictorial representations. To answer this question, performance on the word-problem solution task was compared with performance on the last picture of the picture selection task, because, in both cases, the child had to produce the solution to the problem. In the latter case, however, pictures accompanied the story. If performance differed on the two, this would suggest that the pictures had a beneficial effect. This evidence would not be unequivocal, however, because the picture task always followed the solution task, and mere repeated exposure to the problems could have benefited performance as well. Presentation order was not manipulated in this design, because the major concern was not improving children's performance but identifying their interpretations of the problems and the relationship between interpretation and solution errors. To do this, it was necessary to replicate the standard procedure, that is, observing solution accuracy and error types first, followed by probes of the problems' representations.

With these disclaimers in mind, an analysis of variance was performed on solution accuracy and final picture selection accuracy using problem type (combine 5, compare 4, compare 6, and change 6) and task (word problem and picture selection) as repeated measures. The main effect of task was significant, $F(1, 23) = 8.81$, $MS_e = .08$, $p < .01$. This effect was modified, however, by a Problem Type × Task interaction, $F(3, 69) = 4.64$, $MS_e = .07$, $p < .01$. Simple effects tests performed on this interaction indicated that pictures aided performance on combine-5 and compare-6 problems, $Fs(1, 92) = 10.26$ and 9.17, respectively, $MS_e = .073$, $ps < .01$. Performance on compare 4 was not aided significantly, however, $F(1, 92) = 1.54$, $p > .05$, although differences were in the predicted direction. Performance on change 6, however, actually was hindered by the pictures, $F(1, 92) = 2.02$, $p < .05$. Perhaps the best explanation of this last effect is that children found the picture format used for this problem confusing. The enhanced performance for the other problems, however, suggests that the pictures may have helped children to determine the correct representations of the story situation.

EXPERIMENT 2

The results of Experiment 1 showed a direct relationship between solution success and the ability to conceptualize properly problem situations as described by problem texts. More particularly, breakdowns in interpretations could be traced to the assignment of incorrect interpretations to certain verbal forms. A crucial question is why these misinterpretations occur. According to the linguistic development hypothesis, these misinterpretations occur because children are inexperienced with such forms and have not yet acquired meanings for them. According to the logico-mathematical development view, these misinterpretations indicate a lack of conceptual knowledge concerning part–whole set relations.

The purpose of Experiment 2 was to distinguish between these two alternative explanations. To do this, a sample of children was preselected on the basis of poor performance on combine-5 problems. Combine-5 problems (called combine-2 problems by Riley et al.) were singled out for this comparison because, according to Riley et al., 1983:

> The proposal that children require an understanding of part–whole relations to solve change problems (5) and (6) is supported by Riley's study in which few children in any age group correctly solved these two problem types without first being able to solve combine problems with one of the subsets unknown. . . . In fact, it is this model's lack of understanding of part–whole relations that accounts for its failure on combine 2 problems. The more advanced models have a combine schema that allows them to infer the part–whole relation between, for example, the eight marbles that Joe and Tom have altogether . . . and the five marbles that Joe has. *These relations are not mentioned explicitly in the problem, and without this schema, children simply interpret each line separately* [italics added]. (p. 184)

In other words, combine problems are particularly sparse, providing few cues for choosing a solution strategy. Moreover, these problems are rarely solved correctly before other difficult problems are. According to the logico-mathematical view, children who consistently fail these problems do so because they lack a schema for part–whole relations, and correct performance on these problems can be taken as evidence that part–whole knowledge has been acquired. According to the linguistic development view, these children fail the problem not because they lack a schema for part–whole relations but because they lack a schema mapping the phrases used in the problem (i.e., Mary and John have 5 altogether) onto part–whole knowledge. To say that a child does not have a combine schema from the Riley view means that the child does not understand part–whole relations. To say that a child does not have a combine schema from the Cummins view means that the child does not understand the semantics of the problem text. (To give a simple analogy for this distinction, an English-speaking child may categorize the French word *chien* improperly because he or she does not yet have the concept of *dog* or because he or she does not yet know that *chien* maps onto *dog* in English. The former explanation is analogous to that offered by Riley et al., whereas the latter is analogous to the explanation offered by Cummins et al., for word-problem solving.)

If the logico-mathematical view is correct, children who fail these problems should represent them in ways that do not capture their part–whole structures, regardless of the way the problem is worded. If the linguistic development view is correct, rewording the problem to avoid the joint ownership meaning of conjunctives and the term *altogether* should produce correct depictions of the problem's part–whole structure, whereas standard wordings should produce incorrect depictions.

Method

Subjects

Eleven first graders (7 girls, 4 boys) were chosen from among 23 children from the New Haven Public Schools who were administered the solution and drawing tasks for the combine-5 problem during a preselection process.

These 11 children were chosen because they failed all three instances of the solution task and drew incorrect interpretations. They were middle class and racially mixed. They were tested late in the school year (May).

Materials

The same two number sets used in Experiment 1 were used here to construct word problems. As in Experiment 1, number set assignment was balanced among subjects. For each number set, three standard instances of combine 5 and three reworded instances were constructed. An example of the reworded version follows:

> There are 5 marbles.
> Two of them belong to Mary.
> The rest belong to John.
> How many belong to John?

In addition to these instances, other filler addition and subtraction problems were also constructed for inclusion in the testing session. These were included to ensure that subjects did not simply adopt a subtraction strategy when performing the task. Subjects solved a total of 14 problems: 2 practice problems, 3 standard combine-5 problems, 3 reworded combine-5 problems, and 6 filler problems. The drawing templates described in Experiment 1 were used here for subjects to draw their interpretations and to record their answers to the problems.

Procedure

Subjects were tested in pairs. Problem presentation order was randomized for each pair; number set was also randomly assigned to each pair. Problems were read to the children, and they were required to write their answers on a line next to one of the drawing templates. They were then asked to draw the number of marbles Mary had and the number John had. *Note that the drawing task differed from that used in Experiment 1, where they simply drew a depiction of the first line of one combine-5 problem. Here, they had to draw a depiction of the whole problem situation for each problem they solved.*

The session began with two practice problems, one addition and one subtraction, and subjects were assisted on the drawing task to ensure that they understood the procedure.

Results and Discussion

Proportion drawing and solution accuracy for the standard and reworded combine-5 problems were computed for each subject; mean proportion ac-

TABLE 4
Mean Proportion Drawing and Solution Accuracy for Standard and Reworded Versions of Combine 5 in Experiment 2

	Task	
Version	*Solve*	*Draw*
Standard	.30	.33
Reworded	.85	.85

Note. Means are based on three observations from each of 11 subjects.

curacy is presented in Table 4. An analysis of variance was performed on these data using problem version (standard and reworded) and task (solve and draw) as repeated measures. The analysis returned a single significant result, that of problem version, $F(1, 10) = 28.12$, $MS_e = .11$, $p < .001$. As predicted by the linguistic development hypothesis, these results indicate that children found the reworded versions both easier to represent and easier to solve. It is not clear how the logico-mathematical view can account for these results.

As in Experiment 1, drawing and solution errors were classified. Because children were required to illustrate the entire problem situation rather than just the first line of the problem, it was expected that more errors of a stimulus-matching nature would be observed. This is because, if children misinterpret the first line to mean "each," the second sentence should be interpreted as a contradiction of the first, in that Mary is said to possess a different number of marbles than was stated in the first sentence. As a result, a reasonable strategy would be to assign a new number of marbles to Mary but to retain the originally stated number as belonging to John.

This switch in interpretation strategy was, in fact, observed. In the standard condition, 33% of all drawing responses were stimulus-matching errors, and 15% were "each" interpretation errors. This contrasts with only 6% stimulus-matching drawings and no "each" interpretations in the reworded condition. Similar shifts in strategies were observed on the solution task. If children misinterpret "Mary and John have 5 altogether" as "Mary and John have 5 each," the last line of the problem (i.e., "How many does John have?") should be answered by returning the only number in the problem associated with John (i.e., five). In other words, more given number errors should be produced on standard forms than on reworded forms. This was, in fact, found. Given number errors constituted 46% of all responses in the standard condition but only 9% of all responses in the reworded condition. This pattern implies that given number errors occurred in the standard condition because children were more likely to use a stimulus-matching strategy or "each" interpretation strategy. This inference is supported by the fact that on 60% of trials on which children committed a given number error in the standard condition, they also drew a stimulus-matching depiction of the problem situation; on 33% of these trials, they drew "each" interpretations.

It is interesting to note that De Corte et al. (1985, p. 464) also showed improved performance by rewording this problem type to make the local semantics of the sentences clearer, as follows:

Mary and John have 5 marbles altogether.
Two of these marbles belong to Mary.
The rest belong to John.
How many belong to John?

The improvement in number of correct responses produced by this rewording (43% vs. 57%), however, was not as dramatic as the improvement observed here (30% vs. 85%). This is not surprising considering the way in which children were found to conceptualize the first line of this problem in Experiment 1. If a substantial proportion of De Corte et al.'s subjects interpreted this line to mean "Mary and John *each* have 5 marbles," the phrase in the second line (i.e., "of these") would not necessarily have helped to clarify the problem situation. It may, however, have provided a hint to them to reconsider their interpretation. The rewording used in the present experiment apparently was unambiguous to children, allowing them to map the problem text easily onto tacit knowledge concerning part–whole relations.

GENERAL DISCUSSION

The purpose of these experiments was to obtain clear evidence of the interpretations children assign to verbal forms typically used in word problems. Such information was crucial not only for understanding how children understand such problems but also for testing two conflicting explanations of children's solution errors. The results of Experiment 1 show clearly that children misconceptualize combine and compare problems because they assign incorrect interpretations to the first and second lines, respectively, of those problems. More important, these results show clearly just how children are likely to misinterpret these problems. As argued by Cummins et al. (1988), many children were found to misinterpret "Mary and John have 5 altogether" to mean "Mary and John *each* have five." This misinterpretation led them to construct predictably incoherent problem representations and to choose incorrect solution strategies. Also, as argued by Cummins et al., many children were found to misinterpret comparative verbal forms as statements of possession or to ignore them completely, again leading them to construct incoherent problem representations and to choose incorrect solution strategies. No clear information was obtained concerning their interpretations of change 6, however.

Experiment 2 shows clearly that children's failures to interpret combine problem texts properly were not due to a lack of conceptual knowledge concerning part–whole relations. Children were found to have little difficulty drawing accurate representations of the part–whole structure described by these problems when the texts were reworded in language that was unambiguous. Rewording the first line of this problem as "There are 5 marbles" leaves little room for misinterpretation of the initial problem state. As a result, solution accuracy nearly tripled.

Taken together, these results offer general support for a linguistic explanation of solution error patterns. Although many researchers have noted the importance of text comprehension processes in problem solving (e.g., Kintsch & Greeno, 1985; Riley et al., 1983), the emphasis has been on dealing with the text as a whole, focusing on the interrelations of many complex levels of processing in producing a coherent problem representation. Although such processes are undoubtedly needed and used in order to produce coherent problem representations, the work presented here strongly suggests that sometimes "simpler" explanations of problem-solving performance are overlooked in favor of more complex ones, to the detriment of explaining the available data.

Perhaps the main point of this work is that children should be conceptualized as novices. Like novices in any other domain, they have considerable knowledge at their disposal but have not yet constructed the schematic structures that map out their knowledge in useful ways. Like adult novices, they rely on surface features and key words in building problem representations and engage in a stimulus-matching strategy when faced with incomprehensible problem statements. As they become more familiar with the domain, they build representations of problem types (e.g., combine, compare, change) that constitute interpretations of problem texts and that dictate solution strategies. Examined from this perspective, the developmental model proposed by Riley et al. (1983) depicts not so much the growth of *conceptual* knowledge concerning part–whole relations as the growth of knowledge concerning *problem types* in the domain. Development of such schemata means acquiring coherent mappings from frequently encountered problem texts onto (already possessed) part–whole knowledge. Improvements in problem-solving skill reflect the growth of such mappings. In the approach taken by Cummins et al. (1988), combine, compare, and change schemata are characterized as mappings from language to part–whole knowledge that is acquired through experience. In other words, children come to learn what comparatives and other such phrases mean as a result of exposure to problem-solving episodes that require them to map "How many more does Mary have than John?" onto their knowledge of part–whole relations. To get such problems right, it is necessary to realize that Mary's set can be treated as the "whole" and John's as the "part" in a part–whole superschema. Doing so fosters understanding that phrases containing "more than" are descriptions of comparison situations and not simple assignments. In this way, children acquire interpretations for such verbal phrase forms.

Perhaps the most important point is that we should not underestimate the extent of children's tacit logico-mathematical knowledge. When we probe their conceptual structures with our questions, we should temper our conclusions about their failures with the possibility that they have not understood what information we want from them.

ACKNOWLEDGMENTS

This work was supported by Biomedical Research Support Grant No. S07-RR07015 to the author while at Yale University. I thank Amy Lyons for her invaluable help in conducting these studies.

REFERENCES

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

Carpenter, T. P., Hiebert, J., & Moser, J. M. (1981). The effect of problem structure on first-graders' initial solution processes for simple addition and subtraction problems. *Journal for Research in Mathematics Education, 12,* 27–39.

Cummins, D. D., Kintsch, W., Reusser, K., & Weimer, R. (1988). The role of understanding in solving word problems. *Cognitive Psychology, 20,* 405–438.

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

De Corte, E., & Verschaffel, L. (1986, April). *Eye-movement data as access to solution processes of elementary addition and subtraction problems.* Paper presented at the annual meeting of the American Educational Research Association, San Francisco.

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

De Corte, E., Verschaffel, L., & De Win, L. (1985). The influence of rewording verbal problems on children's problem representation and solutions. *Journal of Educational Psychology, 77,* 460–470.

Fuson, K. C., & Hall, J. W. (1983). The acquisition of early number word meanings: A conceptual analysis and review. In H. P. Ginsberg (Ed.), *The development of mathematical thinking* (pp. 50–109). Orlando, FL: Academic.

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

Inhelder, B., & Piaget, J. (1964). *The early growth of the child.* New York: Harper & Row.

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

Markman, E. M. (1973). The facilitation of part–whole comparisons by use of the collective noun "family." *Child Development, 44,* 837–840.

Markman, E., & Siebert, J. (1976). Classes and collections: Internal organization and resulting holistic properties. *Cognitive Psychology, 8,* 561–577.

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

Nesher, P. (1982). Levels of description in the analysis of addition and subtraction word problems. In T. P. Carpenter, J. M. Moser, & T. A. Romberg (Eds.), *Addition and subtraction: A cognitive perspective* (pp. 25-38). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

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

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

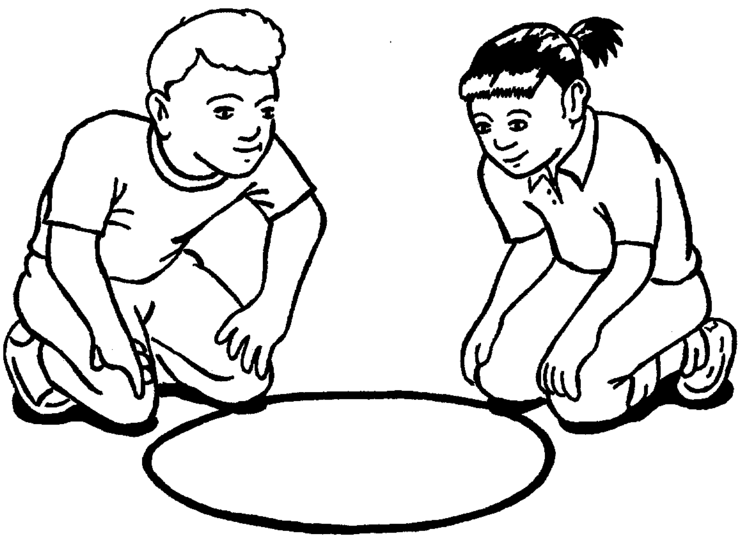
Smith, L. B., & Kemler, D. G. (1978). Levels of experienced dimensionality in children and adults. *Cognitive Psychology, 10,* 502-532.

Trabasso, T., Isen, A. M., Dolecki, P., McLanahan, A. G., Riley, C. A., & Tucker, T. (1978). How do children solve class inclusion problems? In R. S. Siegler (Ed.), *Children's thinking: What develops?* (pp. 151-180). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

APPENDIX

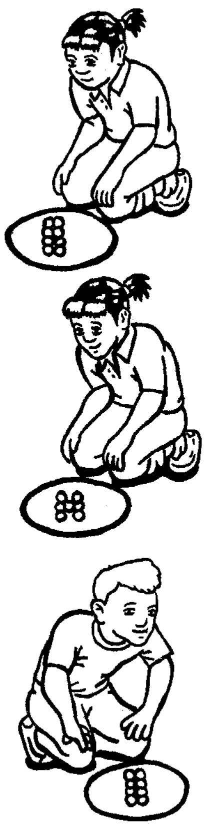
Pictures used in the Picture Selection and Drawing Task

Template for Drawing Task:

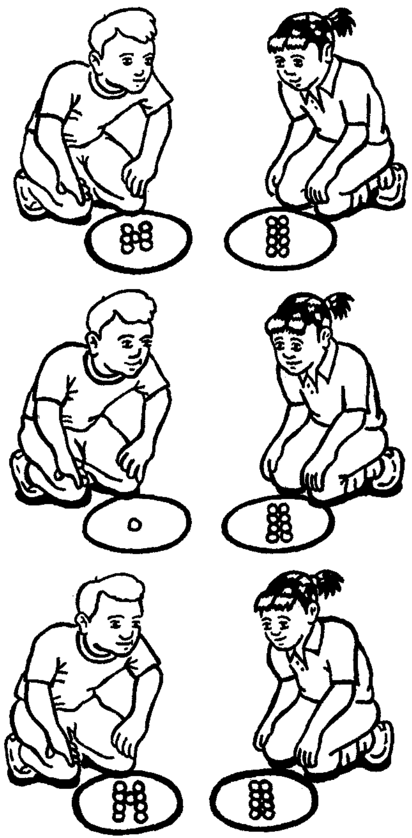


Pictures for Selection Task: See following pages
(Note: When used in the experiments, each vertical group of three was presented on individual pages, and presentation order of targets and foils on each page was randomized.)

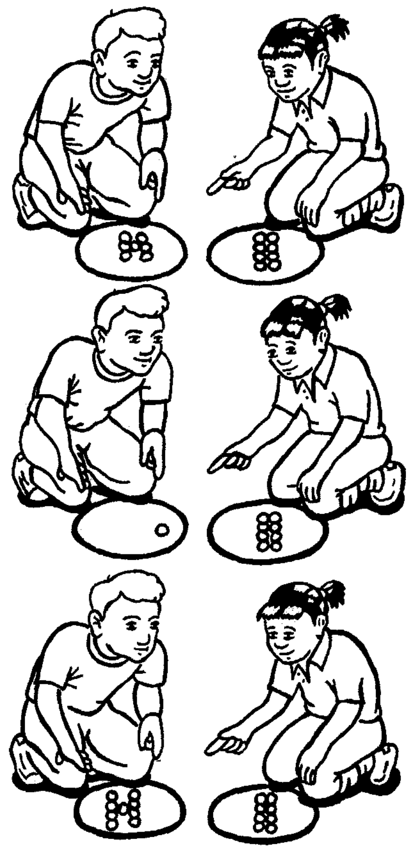
Compare 4
Mary has 8 marbles.



John has 1 marble less than Mary.

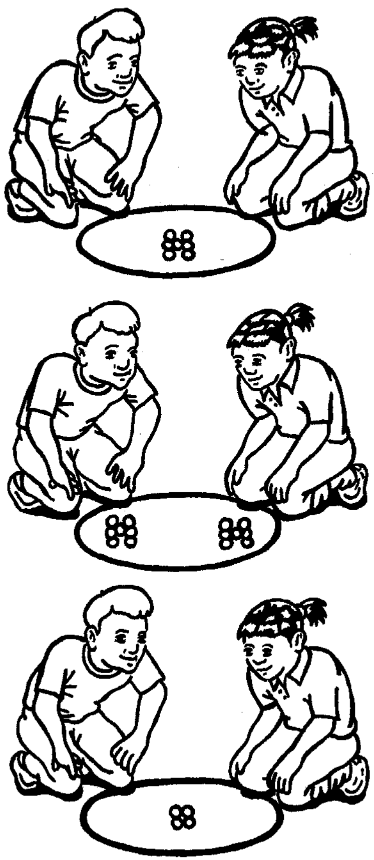


How many does John have?

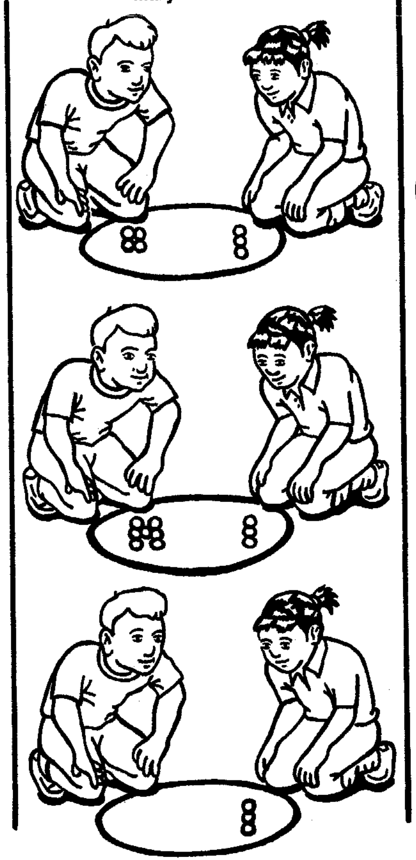


Combine 5

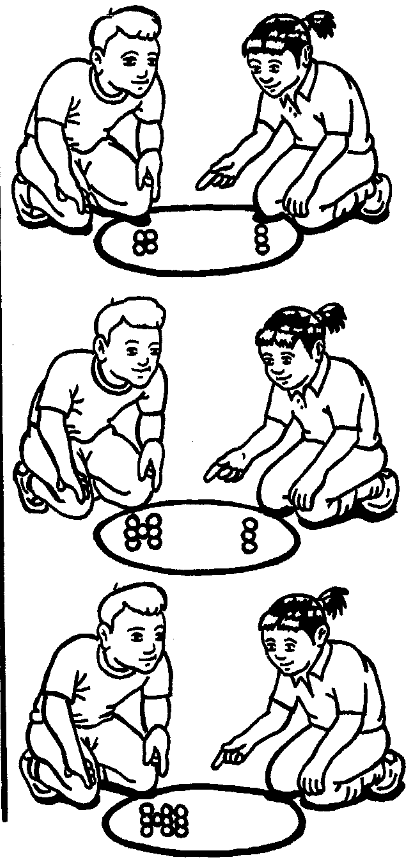
Mary and John have 7 marbles altogether.



Mary has 3 marbles.

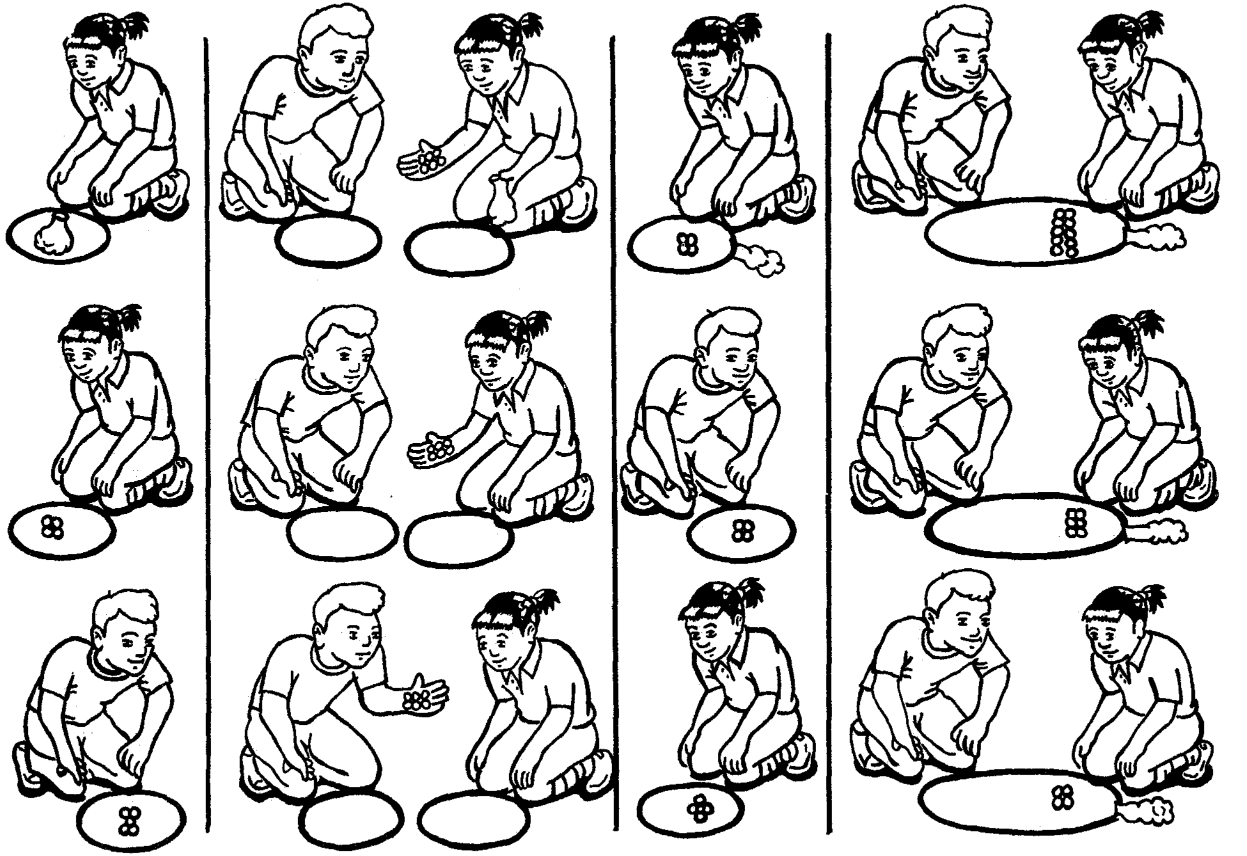


How many does John have?



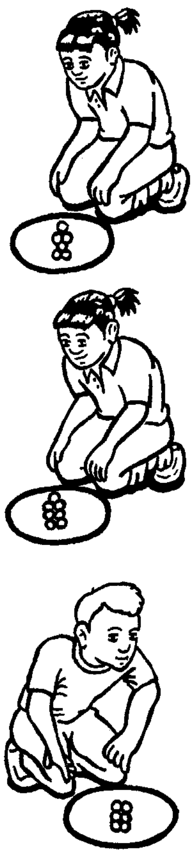
Change 6

Mary has some marbles. Then she gave 6 to John. Now Mary has 4 marbles. How many did she have in the beginning?

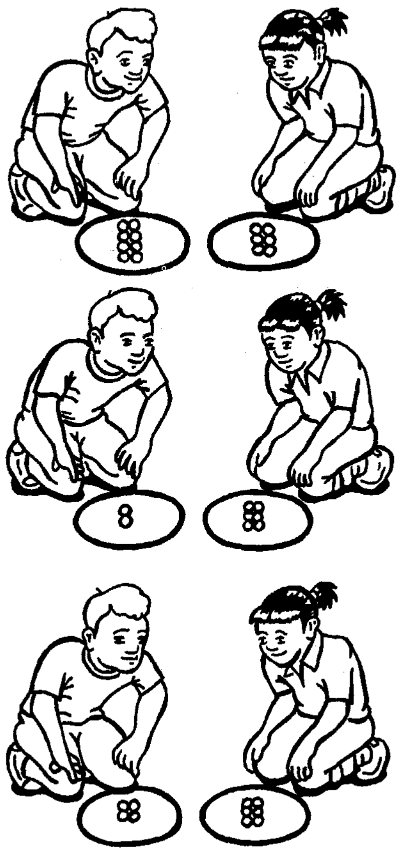


Compare 6

Mary has 6 marbles.



She has 2 marbles less than John.



How many does John have?