COGNITION AND INSTRUCTION, 1985, *2* (1) 41–57

Developmental Differences in Understanding and Solving Simple Mathematics Word Problems

Romelia V. Morales, Valerie J. Shute, and
James W. Pellegrino
University of California
Santa Barbara

The hypothesis was tested that older children, fifth/sixth graders versus third graders, who are more accurate in word problem solution, would show evidence of problem differentiation consistent with schema theories of problem representation and solution. A mathematics word problem sorting task and an accuracy test were used to assess problem representation and accuracy of solution across 16 different problem types. Cluster analyses of sorting data revealed that older children demonstrated sorting patterns consistent with a refined schema theory. Younger children were both less accurate and less systematic in their sorting. Idealized sorting patterns representing schema theories as well as surface structure problem similarity were tested against actual sorting patterns. Regression analyses confirmed that the fifth/sixth graders' and third graders' sorting data were best fit by a schematic sort and by a surface structure sort, respectively. Additionally, an English versus Spanish language contrast showed no effect of language on either accuracy of solution or the sorting patterns of the math word problems.

One of the most problematic areas of the mathematics curriculum involves the solution of word problems. Even though children have mastered basic addition and subtraction operations, they experience considerable difficulty with simple word problems that require application of these procedures. Furthermore, there are systematic differences between and within grades in the accuracy of solving specific types of problems (e.g., Briars & Larkin, 1984; Riley, Greeno, & Heller, 1983; Rosenthal & Resnick, 1974). Interpretation of

Requests for reprints should be sent to James W. Pellegrino, Department of Education, University of Caliornia, Santa Barbara, CA 93106.

such results has focused on the conceptual and procedural knowledge required for problem solution and the importance of conceptual knowledge for initial problem representation. Riley et al. posit that "improvement in performance results mainly from improved understanding of certain conceptual relationships" (p. 154). Vergnaud (1982) also stresses the need for conceptual understanding of word problems for solution because, although certain math word problems may have the same mathematical structure, they can have very different underlying conceptual structures. Consistent with this hypothesis, Carpenter and Moser (1982) found that, in order to understand and solve simple math word problems, certain prerequisite conceptual knowledge is needed. This knowledge includes the understanding of informal concepts such as one-to-one correspondence, class inclusion, and part-whole relationships.

According to Riley et al. (1983), word problems requiring simple addition or subtraction operations can be grouped into four major classes. Table 1 represents each of the four major classes of problems and the specific subtypes within each class. The designations are based upon the classification systems of Riley et al., and Carpenter and Moser (1982). The first class consists of *change* (CH) problems with three specific subtypes, all of which involve an exchange of quantity; differences between subtypes reflect the nature of the unknown: start, change, or result sets. The second class consists of *equalize* (EQ) problems, which are a variant of the change problems and have the phrases "must give away" or "must get," in addition to the idea of comparison and equalization indicated by, "to have as many as." The third problem class consists of the *combine* (CB) problems with two specific subtypes. The combine problems require joining and separating sets, but not by any action explicitly indicated in the word problem. The difference between subtypes is whether the unknown is the total set or one of the subsets. The fourth problem class consists of the *compare* (CP) problems with three specific subtypes. All of these problems include the phrases "more than" or "less than," which indicate the comparison of sets. Differences between subtypes again reflect the nature of the unknown.

The subtypes within three of the four problem classes are arranged in order of difficulty. For both the change (CH) and compare (CP) problems, consisting of three specific subtypes, difficulty is a function of where the unknown is situated. Combine problem CB1, which requires finding the total set, is easier to solve than CB2, which requires finding a subset. The two equalize (EQ) problems are of equivalent difficulty. Several studies have provided data on the relative difficulty of the various subtypes (e.g., Hiebert, 1981; Ibarra & Lindvall, 1979; Lindvall & Ibarra, 1980; Riley et al., 1983). The easiest problems are CH1, CH2, and CB1; the most difficult are CH5, CH6, CB2, CP5, and CP6. The remaining problems are of intermediate lev-

TABLE 1
Riley, Greeno, and Heller Classification of Word Problems

CHANGE (CH)

Result Unknown

1. Pete had 7 marbles.
   Then Sam gave him 5 more marbles.
   How many marbles does Pete have now?

2. Terry had 12 marbles.
   Then she gave 4 marbles to Pat.
   How many marbles does Terry have now?

Change Unknown

3. Allen had 9 marbles.
   Then Ken gave him some more marbles.
   Now Allen has 13 marbles.
   How many marbles did Ken give him?

4. Janet had 14 marbles.
   Then she gave some marbles to Sue.
   Now Janet has 6 marbles.
   How many marbles did she give to Sue?

Start Unknown

5. Emily had some marbles.
   Then Ana gave her 8 more marbles.
   Now Emily has 14 marbles.
   How many marbles did Emily have in the beginning?

6. David had some marbles.
   Then he gave 6 marbles to Jim.
   Now David has 9 marbles.
   How many marbles did David have in the beginning?

EQUALIZE (EQ)

1. Rose has 7 marbles.
   Dora has 10 marbles.
   How many marbles must Rose get to have as many as Dora?

2. Nancy has 6 marbles.
   Eve has 3 marbles.
   How many marbles does Nancy need to give away to have as many as Eve?

COMBINE (CB)

Total Set Unknown

1. Fred has 7 marbles.
   John has 5 marbles.
   How many marbles do they have altogether?

Subset Unknown

2. Eddie and Roy have 11 marbles altogether.
   Eddie has 4 marbles.
   How many marbles does Roy have?

COMPARE (CP)

Difference Unknown

1. Jane has 12 marbles.
   Mary has 7 marbles.
   How many marbles does Jane have more than Mary?

2. Jack has 11 marbles.
   Luis has 3 marbles.
   How many marbles does Luis have less than Jack?

Compared Quantity Unknown

3. Bill has 9 marbles.
   James has 7 more marbles than Bill.
   How many marbles does James have?

4. Joe has 12 marbles.
   Tom has 3 marbles less than Joe.
   How many marbles does Tom have?

Referent Unknown

5. Jerry has 10 marbles.
   He has 4 more marbles than Bob.
   How many marbles does Bob have?

6. Tony has 8 marbles.
   He has 5 marbles less than Henry.
   How many marbles does Henry have?

els of difficulty. Riley et al. have also shown that the relative difficulties of problems are consistent across several grade levels.

According to Riley et al. (1983), successful solution of the different types of problems illustrated in Table 1 is presumed to depend on the availability of conceptual representations or "schemata" corresponding to change, combine, and comparison situations. A schema is an organized structure consisting of elements and relations. The elements are the quantities such as the start, change, and result sets for change problems. The elements are organized in terms of quantitative, temporal, and logical relations defining a general problem class. Given the availability of an appropriate schema in memory, the child must be able to map the verbal problem statement onto the appropriate schema with the appropriate assignment of specific quantities to the slots or elements of the schematic structure. Differences in solution accuracy between problems within a class, for example, CH5 versus CH1, are presumed to result partially from an incomplete schema and/or difficulties in the mapping process.

A related theory of problem solution has been proposed by Briars and Larkin (1984). Their model, CHIPS (*C*oncrete *H*umanlike *I*nferential *P*roblem *S*olver), simulates the psychological processes a child engages in while solving word problems and identifies the degree of difficulty he encounters. The modus operandi of CHIPS involves the formulation of list structures, each representing a counter. It then uses these counters to build a representation of the word problem, ultimately resulting in a schema, a large, organized data structure. When an action is completed, the schema acquires the status "done" and the word problem is solved.

In the Briars and Larkin theory, several elements are believed to contribute to the degree of difficulty in a word problem. Table 2 illustrates the elements involved for each of the problem subtypes illustrated in Table 1. The first element is whether the required action involved in the word problem is cued or not cued. Action-cued problems, group L1, are easier to solve than problems with no action cues. Another aspect is whether the problem requires a single-role or a double-role counter; the former are easier to solve. A third aspect of difficulty in this model is one of rerepresentation. Word problems that need to be rerepresented for solution are more difficult than those not requiring this process. For example, a problem asking, "How much did Joe have in the beginning?" is more difficult than one asking, "How much does Joe have now?" In comparison problems there are additional language factors, such as consistent or conflicting language, which affect solution. Shown at the bottom of Table 2 are predictions about problem difficulty from Briars and Larkin's theory. Within the group of action-cued problems (L1), those with single-role counters are solved more readily than those with double-role counters, which in turn are solved more readily than those requiring rerepresentation. Across groups, action-cued problems (L1) are solved more

TABLE 2
Briars and Larkin Analysis of Word Problem Characteristics

Problem Type	*Characteristics*	*Equation*	*Group*
Change Problems			
CH1 & CH2	Action Cues & Single-Role Counters	$A \pm B = ?$	*L1(SR)*
CH3 & CH4	Action Cues & Double-Role Counters	$A \pm ? = B$	*L1(DR)*
CH5 & CH6	Action Cues & Rerepresentation	$? \pm A = B$	*L1(RR)*
Equalize Problems			
EQ1 & EQ2	Action Cues & Double-Role Counters	$A \pm ? = B$	*L1(DR)*
Combine Problems			
CB1	Implicit Action Cues & Single-Role Counters	$A + B = ?$	*L1(SR)*
CB2	No Action Cues & Single-Role Counters	$A = B + ?$	*L2*
Compare Problems			
CP1 & CP2	No Action Cues & How Many More-Less	$A - B = ?$	*L3*
CP3 & CP4	No Action Cues & Consistent Language	$A \pm B = ?$	*L4*
CP5 & CP6	No Action Cues & Conflicting Language	$A \pm B = ?$	*L5*
Predictions:			
Within *L*1:	*L1(SR) >L1(DR)>L1(RR)*		
Across Groups:	*L1>L2; L1>L3>L4>L5*		

readily than those without action cues (L2–L5), and there is a further ordering of difficulty within comparison problems (L3 > L4 > L5). Predictions from Briars and Larkin's theory are consistent with empirical results reported by Riley et al. (1983) and others.

Both the Riley et al., and the Briars and Larkin theories assume that word problem solution is dependent on several types of knowledge. Most important is conceptual knowledge for problem representation that leads to the appropriate selection of action schemata for solution. Once the situation is understood and the quantity assignments are made, selection of an appropriate operation can be done and solution will be accurate barring computational errors. Older children are more successful in solving word problems because of the availability of appropriate schemata for problem representation and the ability to utilize such schemata to differentiate instances of a problem class and to determine the nature of the unknown. The purpose of the present study was to test the hypothesis that older children, who are more accurate in problem solution, show evidence of problem representation and differentiation consistent with the Riley et al. and Briars and Larkin analyses of problem types and schemata.

Third- and fifth-/sixth-grade children were asked to sort the word problems shown in Table 1 on the basis of similarity. Sorting tasks have been used to assess problem representation and the availability of conceptual knowledge such as different problem schemata (e.g., Chi, Feltovich, & Glaser, 1981; Hinsley, Hayes, & Simon, 1977). Hinsley et al. (1977) used a sorting task in their investigation of the schematic representation of word problems. They concluded that students have schemata for many different problem types, and that these schemata are rapidly applied in the process of problem representation and understanding. The initial representation of a problem in terms of a schema, such as an interest or rate problem, leads to the selection of different formulas appropriate for the problem class. In the present study, it was hypothesized that older children would show more systematic sorting behaviors and that the patterns observed would be consistent with the Riley et al. (1983) and Briars and Larkin (1984) classification of problem types. In addition, it would be expected that the sorting behavior of older and more successful solvers would show a further differentiation representing differences between problems within a class. For example, within change problems a separation would appear for problems involving start, change, or result unknown. These problems are also differentiated on the basis of single- versus double-role counters versus rerepresentation.

A second issue that we investigated was whether input language, English versus Spanish, influences accuracy of solution and sorting patterns among students with similar socioeconomic status and cultural backgrounds. This issue was of interest because of the scarcity of research conducted on the cognitive processes and performance of English- and Spanish-speaking Mexican-Americans. A prevailing attitude is that in bilingual classrooms there is an overuse of Spanish, therefore contributing to low school achievement (Hernandez-Chavez, 1984). If this is the case, then children in bilingual classrooms (Spanish and English) might be expected to perform lower than their monolingual counterparts (English only). However, if children in bilingual classrooms are presented problems in their dominant language (Spanish), then no differences would be expected between bilingual and monolingual children in either accuracy of solution or sorting behaviors. This is because the underlying conceptual representation, in canonical form, should transcend input language differences. Schemata are generally understood to be independent of certain surface structure features of problems such as input language.

METHOD

Subjects

The subjects were 111 third- and fifth-/sixth- grade students from four classrooms. All of the subjects were Mexican-American, low SES students ranging in age from 8 to 9 years for the third graders, and from 10 to 13 years for the fifth/sixth graders. The pool was drawn from two bilingual (Spanish/ English) classrooms (third and fifth/sixth grades), and two monolingual (English) classrooms (third and fifth/sixth grades). There were 61 third graders and 50 fifth/sixth graders.

All of the students were first given a 16-item mathematics word problem test. Twenty-four third graders and 24 fifth/sixth graders were then selected from the pool; 12 students from each classroom. The 12 third-grade Spanish readers consisted of 4 boys and 8 girls; third-grade English readers were 3 boys and 9 girls. Both the Spanish and English readers for the fifth/sixth grades consisted of 4 boys and 8 girls.

The sample of 48 subjects was selected to be representative of the language proficiency (Spanish/English) and solution accuracy levels found in the four classrooms. The Language Assessment Survey and the classroom teacher's judgment were used to determine the language proficiency of each student. In the bilingual classrooms, bilinguality or dominance in Spanish was one criterion for selecting the sample, whereas in the monolingual classrooms, comparable expertise in English was the criterion. The classroom teacher's judgment was also used to verify decisions regarding the language in which the students were to read the word problems.

Procedure

All 111 students were tested as a group in their own classrooms; the test contained the 16 different word problems shown in Table 1. The test was given in their respective native languages, and they were given as much time as was needed to complete it.

Three months later, 48 students were selected to participate in the second phase of the study. A substantial amount of time was allowed to elapse before the sorting task was administered so that the child's attitude toward the task would be one of simply sorting the word problems and not solving them as they had previously done. Each student was given a card deck containing the set of 16 word problems. The students individually sorted the word problems in a separate part of their classroom or outside their classroom on a desk. The students were instructed to first read the 16 word problems aloud; specific instructions were subsequently given regarding the sorting task.

> Please sort the word problems into different groups. Read the first problem, put it down, then read the rest of the cards, one by one, removing the cards that are similar to the first card you took out. You may take out as many cards as you think go together with the first card that you took out. Next, put down another card and do the same thing again. Read each problem and take out the ones that are similar to the one that you have just put down. Are there any changes you would like to make? Do you think some of the problems go together with another group? You can change them if you think so.

After the students sorted the cards into similar groups, the experimenter asked the subjects to explain their reasons for doing so. The subjects were allowed to rearrange the cards if any similarities were noticed during the probing. The students' justifications were recorded on a cassette tape.

Materials

The materials for both the problem-solving accuracy test and the sorting task consisted of the 16 word problems shown in Table 1. Different names and quantities were used in each problem, but the objects of exchange or comparison were not changed. In all cases "marbles" were used as the quantities being exchanged or compared, and the problems were kept at a simple computational level.

For the problem-solving accuracy test, the 16 problems were randomly listed on four pages arranged into four sequences based on a Latin square counterbalancing procedure yielding four different arrangements of problems. An English and a Spanish version of the test were created. The Spanish version was a translation of the English text by four bilingual, Mexican-American teachers. Agreement for a final translation was derived by comparing the four translations and by selecting the words and phrases used most frequently by the four bilingual teachers.

For the sorting task, the 16 word problems were typed on 4 $\times$ 6 in. cards; one problem per card. Again, two sets of cards were generated in both English and Spanish. The card decks contained a random order of problems.

RESULTS AND DISCUSSION

Solution Accuracy

The accuracy of word problem solution was scored for all the children and analyzed in a 2 × 2 ANOVA. There was a main effect of Grade, $F(1,107) = 33.99, p < .001$, with mean percent correct values of 78 and 53 for the fifth/sixth and third graders, respectively. The main effects of language and the grade by language interaction were not significant, F's < 1. A similar analysis was conducted for the subsample of children who performed the sorting task. The main effect of Grade was significant, $F(1,44) = 33.48, p < .001$, with mean percent correct values of 83 and 52 for the fifth/sixth and third graders, respectively. The main effects of language and the grade by language interaction were not significant, F's < 1. Because there were no main effects or interactions involving language, in all subsequent analyses language contrasts will not be presented.

Table 3 shows the data for solution accuracy as a function of major problem types. For all problems except CB1, the fifth/sixth graders were more accurate than the third graders, z's > 1.96, p's < .05. For fifth/sixth graders, the easiest problems were CH1 and CH2, EQ1 and EQ2, CB1, and CP1 and CP2. The most difficult problems were CH5 and CH6, and CP5 and CP6. All other problems were intermediate in solution difficulty. This pattern is consistent with data reported by Riley et al. (1983). The data for the third

TABLE 3
Summary of Performance on Problem Types

		Proportion Correct		*Proportion of Errors: Conceptually Based*	
		Grade 3	*Grade 5/6*	*Grade 3*	*Grade 5/6*
Change	1 & 2	.63	.88	.91	.92
	3 & 4	.51	.78	.88	.77
	5 & 6	.49	.67	.84	.91
Equalize	1 & 2	.51	.92	.97	.75
Combine	1*	.87	.94	.88	.67
	2	.43	.72	.91	.86
Compare	1 & 2	.54	.86	.86	.79
	3 & 4	.52	.74	.92	1.00
	5 & 6	.36	.60	.92	.95

*No significant grade difference ($p > .05$)

graders are also consistent with accuracy patterns observed by Riley et al. for younger age groups. Performance was best on CB1 and CH1 and CH2. The worst performance was observed on CP5 and CP6, CB2, and CH5 and CH6. The third graders also had considerable difficulty with the easiest compare problems CP1 and CP2 and with the EQ problems, in contrast to the fifth/sixth graders.

Analyses were done of differences in solution accuracy among problem types within each grade. The contrasts of interest were based on the predictions shown earlier (Table 2), which were derived from the Briars and Larkin (1984) theory. Within the third grade, results were $L1(SR) > L1(DR) = L1(RR)$ and $L1 > L2$, $L1 = L3 = L4 > L5$. The respective z scores for these contrasts were: 4.17, .36, 2.09, .60, .31, and 2.52, with $z > 1.96$, $p < .05$. For fifth/sixth graders the results were $L1(SR) = L1(DR) > L1(RR)$ and $L1 > L2$, $L1 = L3 > L4 > L5$. The respective z scores for these contrasts were: 1.38, 3.61, 1.97, −.73, 2.12, and 2.11, with $z > 1.96$, $p < .05$. Although neither grade produced a pattern of results completely in agreement with the Briars and Larkin predictions, neither grade produced results contradictory to the predictions in terms of condition reversals. In essence, the patterns produced within each grade showed that certain factors that should affect solution did not, and this can be attributed to high overall levels of performance (fifth/sixth) or low overall levels of performance (third). If the data from both grades are averaged, then the differences among problems are almost perfectly in accord with the Briars and Larkin predictions with the one exception being $L1 = L3$. As noted earlier, the data are strongly in accord with predictions and data reported by Riley et al. (1983).

An analysis of errors was performed to determine whether mistakes could be attributed to the selection of the wrong mathematical operation, a conceptual error, versus selection of the right operation but making a computational error. Table 3 shows that errors were almost totally attributable to conceptual errors. Problems were missed because the children did not know which operation was appropriate. This is presumably due to a failure in problem representation and/or selection of an action schema. Children seldom made errors of the type where one of the quantities stated in the original problem was also given as the answer. Although such errors are predicted by the Briars and Larkin theory, we suspect that they did not occur because of demand characteristics of the testing situation and the age of the children. We suspect that if the child did not "understand" the problem, he or she chose an operation consistent with a surface feature of the problem; for example, more or less. In essence, they understood that the quantities in the problem needed to be used to derive an answer, so an operation was performed on those quantities even though it was conceptually inappropriate.

There are several important things to note about the accuracy results. First, the overall accuracy levels are lower than those reported previously for children in the same grades (Riley et al., 1983). This can be attributed to the group-administered format of presentation, the lack of manipulatable objects to aid representation and solution, and perhaps the educational history of the children in this sample. However, the second thing to note is that the patterns observed over problem types are consistent with patterns shown previously for other age groups under more advantageous testing conditions (Riley et al.), as well as predictions from the Briars and Larkin theory. Finally, errors were conceptually rather than computationally based, consistent with the hypothesis that children have difficulty in understanding and representing certain specific problem types.

Sorting Behavior

Exploratory analyses. Hierarchical cluster analyses were performed on the sorting data for the third- and fifth-/sixth-grade children. The language factor was evaluated in preliminary analyses and no differences in sorting patterns attributable to language were obtained, similar to the results for problem solution accuracy. Furthermore, by collapsing across language, greater stability was achieved in the cluster analysis solutions. Composite co-occurrence matrices were created for each grade level. The entries in the matrices represented the total number of children within a grade who sorted a specific problem (e.g., CH1) with another specific problem (e.g., CH2).

The cluster analysis results are shown in Figure 1. Individual items or groups of items that "cluster" together are connected in the figure. They are connected at values reflecting their similarity to each other. For fifth/sixth graders the CH1 and CH2 problems are clustered together at a relatively high similarity value of .75. This pair of problems is subsequently joined with two other pairs (CH3 & CH4, CH5 & CH6) at a relatively low similarity value of .22. Thus, there is one large group of six problems representing three "weakly" related pairs but with "strong" item relationships within each pair. As can be observed in Figure 1, there are both similarities and differences between the grades in the problem groupings. First, each grade shows three large groups and composition of the groups partially overlaps. Second, the third-grade children were less systematic in their sorting, and the maximal similarity value for any pair of problems was .54. This is in contrast to the greater systematicity for the fifth/sixth graders who had a maximal similarity of .75. In general, higher similarity values were obtained for problem clusters in the fifth-/sixth-grade sample.

Of particular interest are the partitions that appeared in the third-grade data. Their groups consisted of CH1 and CH2, CH3 and CH4, CH5 and CH6, EQ1 and CP1 (which at higher levels joined with CB1, CP2, and EQ2), CP4 and CP6, and CP3 and CP5, with the latter two groups merging at a higher level. Combine 2 was an outlier and was subsequently merged with the

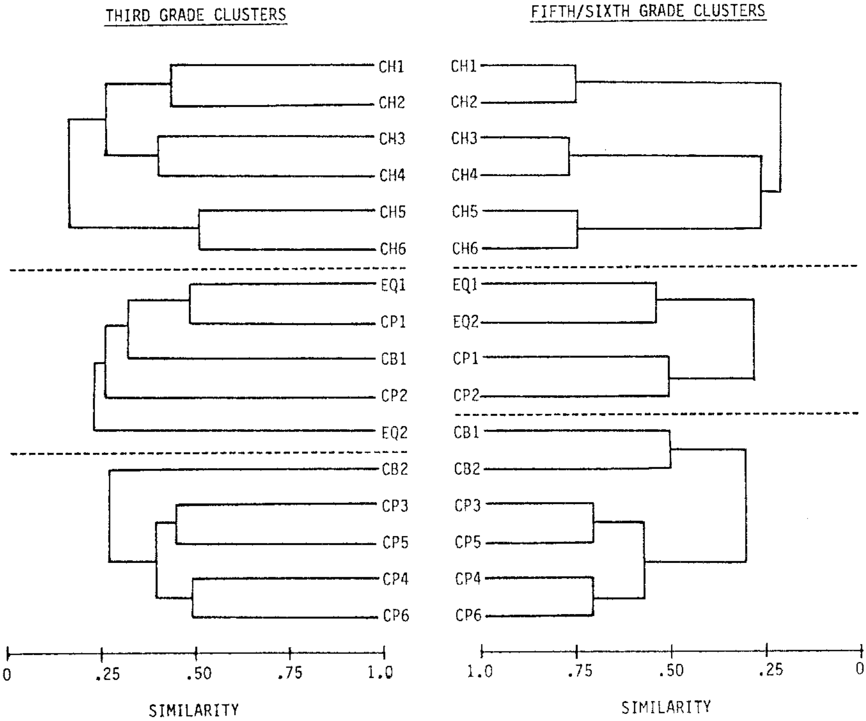

Figure 1 Hierarchical cluster analysis representation of sorting patterns for third- and fifth-/sixth-grade children.

cluster of CP3, CP4, CP5, and CP6. Only the change problems show a sorting pattern consistent with a refined schema theory. More specifically, these problems show grouping based upon an overall problem class as well as a specific unknown within each class of problems. The more difficult compare problems show grouping based upon overall problem class, but the pairing of CP3 with CP5, and CP4 with CP6 is not consistent with specific unknowns; that is, referent versus compared quantity unknown. Instead, these subgroups reflect sorting based upon problem type and a key word in the problem statement, namely the use of *more* (CP4 and CP6) or *less* (CP3 and CP5).

Use of surface structure similarity in sorting is most evident in the large heterogeneous group of EQ1, CP1, CB1, CP2, and EQ2. Each of these problems has two initial quantity assignments of the form: *X* has *n* objects; *Y* has *m* objects, where *X* and *Y* stand for different persons, and *n* and *m* stand for different quantities. Although different operations are required to solve these problems and the schemata vary, they share a similar surface structure and thus appear to be grouped together on this basis. In contrast, CB2, an obvious outlier, has a surface structure unlike any other problem in the entire set and thus tends to be sorted inconsistently. The third-grade data therefore indicate some sorting based upon major problem types and subtypes, but with inconsistencies and evidence of sorting based on surface features that do not bear on problem types or subtypes. This will be elaborated in the following section discussing confirmatory analyses.

The fifth/sixth graders' data yielded more consistent partitions and showed evidence of a more sophisticated sorting strategy based upon specific problem subtypes; that is, sorting by type of unknown within problem type. The clusters were CH1 and CH2, CH3 and CH4, CH5 and CH6, EQ1 and EQ2, CB1 and CB2, CP1 and CP2, CP3 and CP5, and CP4 and CP6. Higher level groups merged the change problems together, the equalize problems with the simpler compare problems, and the more difficult compare problems with the combine problems. Of particular interest is the higher level group consisting of the equalize and simpler compare problems. All four of these problems involve two initial quantity assignments and an implicit or explicit comparison of quantities to determine the difference. It should also be noted that the fifth/sixth graders sorted the more difficult compare problems together, but the individual pairs also reflect attention to the surface structure feature of "more" or "less" in the problem statement. Generally, the fifth-/sixth-grade data show sorting patterns reflecting specific problem subtypes with higher level groups consistent with general problem types or schemata.

Confirmatory analyses. Cluster analysis is an exploratory analysis technique and as such it is insufficient to test a specific hypothesis about the sorting patterns (see Pellegrino & Hubert, 1982). Confirmatory analyses were conducted to compare the co-occurrence matrices for each grade with hypothetical sorts reflecting different theories of the sorting strategy. Three such hypothetical sorts were evaluated. The first of these was based on the Riley et al. (1983) theory and involves sorting by schematic problem type (i.e., change, equalize, combine, and compare), as well as differentiating within change or compare problem types. This hypothetical sort assumes a refined sorting strategy in which problems will be grouped together by specific unknowns.

The second hypothesis that was evaluated was derived from the Briars and Larkin (1984) theory. By differentiating problem types according to whether they contained action cues (explicit, implicit, or no cues), type of role counter (single vs. double, or rerepresentation), and/or language type (consistent vs. conflicting), we were able to construct probable values for the co-occurrence of pairwise problem types in the sorting task. The third hypothesis that was evaluated was sorting according to surface structure characteristics of the problem types. Variables such as the number of sentences in the problem and matching sentence (phrase) structure were used to quantify and match problem types.

Matrices representing each of the three sorting hypotheses were created and then normalized, as were the actual co-occurrence matrices for each grade. Table 4 contains results from correlation analyses where each hypothesis matrix was correlated with the other hypothesis matrices and with the actual sorting patterns in each grade. Matrices derived from each of the two theories of problem representation are highly correlated, as one might expect, and both show an intermediate level of correlation with the matrix representing surface structure problem similarity. Of greater interest are the correlations obtained for each hypothesis matrix with actual sorting patterns. For the third graders, the highest correlation is obtained for the surface structure matrix. In contrast, for the fifth/sixth graders the highest correlation is obtained for the matrix based on the Riley et al. theory of problem types. The other hypothesis matrices show slightly lower correlations with the fifth-/sixth-grade sorting data. The simple correlations shown in Table 4 suggest that the third graders' sorting is best explained in terms of surface structure problem features, while the fifth/sixth graders' sorting data is best explained in terms of conceptual and surface structure problem features.

To further test such an interpretation, multiple regression analyses were conducted. The predictors were each of the hypothesis matrices and the criterion was actual sorting behavior within each grade. For the third grade, there was no significant increase in variance accounted for when either conceptual sorting matrix was added to the prediction derived from the surface structure matrix alone. For the fifth/sixth grade, the optimal prediction was obtained from the combination of the Riley et al. conceptual sorting matrix and the surface structure matrix, $r = .73$. Both predictors were significant contributors, and the conceptual sorting matrix made more of a contribution to the prediction of sorting performance. The Briars and Larkin conceptual sorting

TABLE 4
Intercorrelation Matrix for Hypothesized and Actual Sorting Patterns*

	BL	*SS*	*3*	*5/6*
Conceptual Sorting based on Riley, Greeno, and Heller (RGH)	.81	.57	.38	.68
Conceptual Sorting based on Briars and Larkin (BL)	...	.48	.30	.54
Sorting based on Surface Structure Features (SS)	...	...	.58	.60
Third-Grade Sorts (3)	...	...	...	.62
Fifth/Sixth-Grade Sorts (5/6)	...	...	...	...

*All p's < .01

matrix failed to contribute any additional variance to the regression equation, due to the large amount of similarity and shared variance between the two conceptual sorting hypotheses.

Individual Subject Data

Similar analyses of sorting patterns were conducted for each child. Correlations were computed between the individual students' sorting patterns and two hypothesized sorting patterns: Riley et al. and surface structure. Sorting based on the Briars and Larkin theory was highly correlated with sorting based on the Riley et al. theory, and therefore was not used. Within each grade four possibilities existed for each child's sorting pattern: (a) evidence of conceptual sorting only, (b) evidence of surface structure sorting only, (c) evidence of both conceptual and surface structure sorting, and (d) no evidence of either type of sorting. Within the third grade the percentage distribution was as follows: 0% conceptual only, 21% surface structure only, 38% conceptual and surface structure, and 42% neither type of sorting. Thus, only 38% sorted conceptually while 59% showed evidence of surface structure sorting. For the fifth/sixth graders the percentage distribution was as follows: 13% conceptual only, 4% surface structure only, 71% conceptual and surface structure, and 13% neither type of sorting. Compared to the third graders, the fifth/sixth graders had a greater absolute and relative percentage of individuals sorting conceptually versus syntactically, 84% versus 75%. These findings for individual children support those found at the group level.

Correlations were computed between those students sorting "with structure" (i.e., individuals showing sorting strategies consistent with either the Riley et al. or surface structure representation of problem types) and accuracy of solution. The third graders who sorted conceptually confirmed the Riley et al. contention that conceptual understanding results in better performance. The correlation between accuracy and level of conceptual sorting among the third graders was highly significant: $r = .81$; $p < .005$. Contrasting with this, the third graders who sorted on the basis of surface structure did not have a significant correlation between accuracy and level of surface structure sorting, as predicted. A similar analysis for the fifth/sixth grade did not show any significant correlations between accuracy scores and conceptual or surface structure sorting. This is due in part to the high level of overall accuracy for the fifth/sixth graders and, hence, the low variability.

GENERAL DISCUSSION

The present study was designed to examine several issues regarding the processing and solution of simple word problems. First, it replicates previous results (e.g., Hiebert, 1981; Lindvall & Ibarra, 1980; Riley, et al., 1983) showing that older children are more successful in solving simple word problems, and that the differences between and within grades are a systematic function of problem types. Evidence was also provided that errors in solution were primarily a function of selection of the appropriate procedure; that is, conceptually based rather than calculational. This would be expected given models of performance that assume that errors result from inappropriate problem representation due to the lack of conceptual knowledge and schemata for problem understanding.

Second, differences between and within grades in the solution of problem types are not related to language differences. Within each grade there were no differences between solution rates or patterns as a function of presentation in English versus Spanish for children with matching language dominance. The lack of a language effect was also observed in the sorting behavior within each grade. These results suggest that previous research and theory on the solution of word problems is generalizable to other language and socioeconomic groups. This might be expected, since the primary assumption about problem solution is that it depends on the availability of conceptual knowledge and schemata that are language independent and relatively simple.

The major purpose of the present study was to test the hypothesis that older children who are more accurate in problem solution have refined conceptual schemata corresponding to the major problem classes described by Riley et al. (1983) and Briars and Larkin (1984). Problem-sorting patterns were examined and shown to differ between the third- and fifth-/sixth-grade children. In particular, the older children showed a sorting pattern that not only differentiated each of the major problem classes, but also differentiated subtypes within classes reflecting specific types of unknowns (Riley et al.). This differentiation also corresponds to features relevant for solution in the CHIPS model (Briars & Larkin). The highly refined differentiation was observed for all problem classes. In contrast, the younger children showed a less systematic sorting pattern, although there was some correspondence to the major problem classes. In the younger age group, differentiation by unknown was only apparent for the change problems. The hypothesized sorting pattern that best fit the data of younger children was one of surface structure characteristics of problems. Neither age group showed any evidence of sorting by arithmetic operation required for solution. Thus, the data indicate that older children are able to precisely classify and group word problems that represent similar specific instances of more general schemata, while the younger children attend more to surface characteristics of the problems.

With respect to theories of knowledge and problem solving, the present results are consistent with the assumption that conceptual knowledge and schemata must be available if individuals are to represent specific problem types and then select appropriate procedures for problem solution. Thus, the results are similar to those obtained by Hinsley et al. (1977) for word problems. The older children, as a group, sorted schematically and also were more accurate than the younger children in actual problem solution. Within the younger group, children who showed higher levels of schematic sorting also had higher levels of problem solution accuracy. Given the relationship between conceptual knowledge for problem representation and solution accuracy, an important instructional issue is whether it is possible to directly or indirectly teach such conceptual knowledge.

REFERENCES

Briars, D. J., & Larkin, J. H. (1984). An integrated model of skill in solving elementary word problems. *Cognition and Instruction, 1*(3), 245–296.

Carpenter, T. P., & Moser, J. M. (1982). The development of addition and subtraction problem-solving skills. In T. P. Carpenter, J. M. Moser, & T. Romberg (Eds.), *Addition and subtraction: A cognitive perspective* (pp. 9–24). Hillsdale, NJ: Lawrence Erlbaum Associates.

Chi, M. T. H., Feltovich, P. J., & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive Science, 5,* 121–152.

Hernandez-Chavez, E. (1984). The inadequacy of English immersion education as an educational approach for language minority students in the United States. In *Studies on Immersion education: A collection for United States educators* (pp. 144–183). Sacramento, CA: California State Department of Education, Office of Bilingual Bicultural Education.

Hiebert, J. (1981). *Young children's solution process for verbal addition and subtraction problems: The effect of the position of the unknown set.* Paper presented at the 59th Annual Meeting of the National Council of Teachers of Mathematics, St. Louis, MO.

Hinsley, D., Hayes, J. R., & Simon, H. (1977). From words to equations: Meaning and representation in algebra word problems. In M. A. Just & P. A. Carpenter (Eds.), *Cognitive process in comprehension* (pp. 89–106). Hillsdale, NJ: Lawrence Erlbaum Associates.

Ibarra, C. G., & Lindvall, C. M. (1979). *An investigation of factors associated with children's comprehension of simple story problems involving addition and subtraction prior to formal instruction on these operations.* Paper presented at the 57th Annual Meeting of the National Council of Teachers of Mathematics, Boston, MA.

Lindvall, C. M., & Ibarra, C. G. (1980). Incorrect procedures used by primary grade pupils in solving open addition and subtraction sentences. *Journal for Research in Mathematics Education, 11*(1), 50–62.

Pellegrino, J. W., & Hubert, L. J. (1982). The analysis of organization and structure in free recall. In C. R. Puff (Ed.), *Handbook of research methods in human memory and cognition* (pp. 129–172). New York: Academic Press.

Riley, M., Greeno, J., & Heller, J. (1983). Development of children's problem-solving ability in arithmetic. In H. Ginsburg (Ed.), *The development of mathematical thinking* (pp. 153–196). New York: Academic Press.

Rosenthal, D. J. A., & Resnick, L. B. (1974). Children's solution processes in arithmetic word problems. *Journal of Educational Psychology, 66,* 817–825.

Vergnaud, G. (1982). A classification of cognitive tasks and operations of thought involved in addition and subtraction problems. In T. P. Carpenter, J. M. Moser, & T. Romberg (Eds.), *Addition and subtraction: A cognitive perspective* (pp. 39–59). Hillsdale, NJ: Lawrence Erlbaum Associates.