COGNITION AND INSTRUCTION, 1987, *4*(3), 187-202

Arithmetic Problems Formulation and Working Memory Load

Michel Fayol, Hervé Abdi, and
Jean-Emile Gombert
Laboratoire de Psychologie
Université de Bourgogne

First, third, and fifth graders (French children in American-numbered grades) were asked to solve arithmetic problems in which an initial state was modified by two successive transformations. Three independent variables were manipulated systematically. First, the unknown state was either the final state (S1) or the initial state (S2). Second, either the known state (O1) or the transformations (O2) appeared in the first place in the problem wording. Third, the question was either located at the end (Q1) or at the beginning (Q2) of the problem text. As anticipated, these modifications strongly affected the performances at every age: S1 appears clearly easier than S2; O1 leads to a better performance than O2; and Q1 is better than Q2. The third graders participated in a second experiment in which they had to solve the same problems but with easier numbers. As in the first experiment, we found strong effects related to the problem types (S1 vs. S2) and to the place of the transformations in the problem. However, modifying the place of the question did not show any reliable effect. The theoretical implications of these results are discussed in terms of span of working memory.

Arriving at the solution of a verbally presented arithmetic problem implies the use of at least three kinds of "mental operations." First, we must store the information as a whole until we know the task. Second, we must find an appropriate schemata for organizing and solving the kind of problem encountered in long-term memory (LTM).[1] Third, we must apply the

Requests for reprints should be sent to Michel Fayol, Laboratoire de Psychologie, Ancienne Faculté, Université de Bourgogne, 36 Rue Chabot-Charny, 21000 Dijon, France.

[1]Another possibility is that a structure may be constructed by a comprehension procedure that could combine components of information in an appropriate way. Nevertheless, such a procedure also implies (a) a storage of information and (b) a search in LTM of relevant (partial) schemata.

problem solving process to the data at hand and control its execution. To this end, we must rely on computations or on prestored knowledge in LTM (e.g., "number facts" such as $3 + 2 = 5$).

According to Baddeley and Hitch's (1974) formulation (see also Baddeley, 1986; Hitch, 1980), two different types of components make up the working memory (WM) system. The first component is a *control executive processor:* a kind of high-level, multipurpose system, initiating, monitoring, and coordinating some lower level processes. The second component consists of some *slave systems,* dealing with short-term storage and treatment of specialized information. These slave systems include the *articulatory loop,* dealing with storage of phonemic information, and the *visuospatial scratch pad,* dealing with visuospatial information. We expect interferences to occur within such a limited "work-space," when the storage load increases in a slave system (e.g., the articulatory loop), the performance level on a concurrent task (e.g., reasoning) should decrease, and/or the time needed to perform the task should increase. Thus *work capacity* can be allocated, at least partially, either to store or to process the information according to the demands of the task at hand.

The developmental point of view adopted here increases the complexity of the problem. As is well known, ontogenetic development implies two kinds of progress. On the one hand, children acquire more facts, algorithms, and schemata which are stored in LTM (Naus, 1982). On the other hand, the span of the WM system seems to increase or to be used more efficiently with age (cf. Case, 1982; Case, Kurland, & Goldberg, 1982; Dempster, 1981; Hitch, 1978; Hitch & Halliday, 1983; Hulme, Thompson, Muir, & Lawrence, 1984; Pascual-Leone, 1970).

In line with that broad theoretical approach, we are studying verbally presented arithmetic problems. We take as our working hypothesis that children do not have well-organized schemata for different types of verbally presented problems. They must therefore build a representation in a "bottom-up" manner from the text itself; this heavily taxes WM. Thus, computational difficulties compete with building a representation for WM capacity. Consequently, children will be more dependent on the specific wording of problem statements than will adults (cf. Hudson, 1983; Riley, Greeno, & Heller, 1983; Vergnaud, 1982).

Accordingly, we selected two types of problems, differing in their semantic structure. All problems have the same underlying "change pattern" (Riley et al., 1983): An initial state (I) is modified by two successive "transformations" ($T1$ and $T2$ such that $T1 + T2 = T$), leading to a final state (F). We systematically varied the unknown state either for F (problem structure S1) or for I (problem structure S2). In doing so, we contrasted two types of "change problem": those with *result unknown* versus those with *start unknown* (Riley et al., 1983). (See Table 1 for an example.) In line with

TABLE 1
Examples of Problem Types

Semantic Structure		Examples
S1	Known : *I* (initial state) *T1* and *T2* (transformations) Unknown : *F* (final state)	Paul has ______ candies. His mother gave him ______ candies. How many candies does Paul have now?
S2	Known : *F* (final state) *T1* and *T2* (transformations) Unknown : *I* (initial state)	Paul has now ______ candies. His mother gave him ______ candies. His sister gave him ______ candies. How many candies did Paul have before his mother and his sister gave him candies?

the results of several previous studies, we expected S2 problems to be harder to solve than S1 problems. However, the differences between S2 and S1 should decrease as the age of the children increases.

Moreover, we hypothesized that the problem's semantic structure alone is not sufficient to trigger the solution process. Some modifications of the formulation can facilitate or inhibit data processing, and thus finding the solution. Two main factors can be singled out:

1. The location of the question (Q) at the end of the problem (Q1) or at the beginning (Q2). An early question could act like an "advanced organizer" allowing the subject to search and activate the relevant (or supposed to be relevant) schema of representation and solution in LTM. The data can then be organized quickly and processed in a somewhat "top-down" manner. Consequently, the working memory load is lightened and Q2 will be easier than Q1 (assuming that the correct schemata have been activated in LTM). This hypothesis is consistent with the literature on the comprehension of narrative and expository text (Kieras, 1980; Kosminsky, 1977).[2]
2. The order of information presentation. Either the *state* (O1) or the *transformation* (O2) can appear in the first place. When the transformations are given first, it is possible to avoid storing two transformations by putting the result (*T*) in the place of *T1* and *T2*. Consequently, O2 should facilitate the solution process.

To sum up, three experimental factors emerge from the analysis: the type of problem (S1 vs. S2), the location of the question (Q1 vs. Q2), and the

[2]An alternative possibility suggested by Greeno might be that knowing the goal from the beginning could permit the data following to be processed more efficiently. Such a possibility leads to the same predictions as our hypothesis and seems both compatible and consistent with our theoretical analysis.

order of presentation of the information (O1 vs. O2). These three between-subjects factors were crossed in 2 × 2 × 2 experimental design.

However, a major problem still remains: There exists a clear and well-known age difference in numerical ability. Generally, 6-year-old children solve one-digit addition (e.g., 3 + 4) by counting explicitly. Ten-year-olds, however, have access to LTM stored solutions (Ashcraft, 1982; Ashcraft & Battaglia, 1978; Ashcraft & Fierman, 1982; Groen & Parkman, 1972; Svenson, 1975). Thus, giving the same numerical data at all ages places the younger participants in a state of "WM overload." One possible way to overcome this problem is to equate roughly the level of numerical difficulty of the task for the different age groups. To do so, we pretested several series of three-term additions with several groups of children (ages 6, 8, and 10 years). Participants were tested individually. Additions were selected to match the children's school level (according to textbooks). Each child was asked to perform 20 additions (first graders) or 30 additions (third and fifth graders). Additions were randomized before each presentation. Participants were given 10 sec to give an answer. Time was measured with a stopwatch. The first graders ($n = 10$) did addition of three one-digit numbers without carrying. The third graders ($n = 10$) did addition of one two-digit numbers and two one-digit numbers without carrying (see Table 2 for an example). The fifth graders ($n = 10$) did additions of three two-digit numbers, two ending with 5, and one ending with 0. Only the operations solved by 80% of the participants at a given age were included as stimuli in the experiment. The children were allowed to count aloud or to use their fingers. Although many different number series were tested with several groups of children for each age, this was not, strictly speaking, a controlled experiment. Our aim, however, was to obtain a rough control for the differential WM load by age. We used that procedure only for addition. Table 2 shows some examples of the number series used for different age levels.

We addressed the problem of the impact of numerical data magnitude more precisely in the second experiment. In this experiment, we presented the problems with two different series of numbers. The ones used in the first experiment we called "difficult numbers"; these are abbreviated N2. The

TABLE 2
Examples of Number Series Used in Experiment 1

	First Grade	*Third Grade*	*Fifth Grade*
Initial state	2	24	25
Transformation *T1*	4	3	40
Transformation *T2*	1	5	15
Final state	7	32	80

second series was constructed to reduce the "computational load" and are called "easy numbers" and denoted by N1. See Table 2 for an example. We hypothesized that solving the N1 problems should tax the WM work-space less, particularly because the "central executive" that initiates, controls, and coordinates the computations has a lighter load for N1 problems than for N2 problems. Consequently, we expected the number of miscalculations, errors, and semantic mistakes to decrease for N1 problems. Obviously, because working memory space is limited, every component taking up a large part of it reduces the remaining space available for other activities. Thus, increasing the computational difficulties should lead to a "lowering of reasoning" about the "semantic interpretation."

EXPERIMENT 1

Method

Three between-subjects variables were manipulated. First, the unknown state was either the final state (S1 problems) or the initial state (S2 problems). Second, the order of information presentation was counterbalanced: either state first (O1) or transformation first (O2). Third, the question was either located at the end (Q1) or at the beginning (Q2) of the problem text. These three factors were crossed in a 2 × 2 × 2 factorial design. The problems were constructed with eight varieties of "countable" objects: marbles, candies, dishes, pencils, sheets, books, persons, and cars. For each age level, every object was associated with one and only one numerical series (see Table 3 for some examples). These series were selected to offer roughly the same difficulty level at each age.

Children. Sixty-four children in each of three age groups participated in the first experiment (mean ages: 6.8 years, 8.8 years, and 10.7 years). The children came from three predominantly upper-middle-class schools. They were tested individually in sessions lasting about 40 min.

Procedure. In the test sessions, children were first given a series of warm-up exercises. These exercises differed for each age group. Six-year-olds were asked to count aloud forward and backward from 1 to 10. Eight-year-olds were asked to count forward from 10 to 40 by steps of 2, and then to count backward from 40 to 10 by steps of 5. Ten-year-olds were asked to count, by steps of 5, from 20 to 90, and then backward from 100 to 10. After these exercises, participants were told to count objects shown to them on cards without touching them. They were presented with a series of five cards with a number of dots varying from 3 to 7 drawn on them (the

TABLE 3
Examples of Every Modality of the Experimental Problems

S1 O1 Q1	Paul had ______ candies. His mother gave him ______ candies. His sister gave him ______ candies. How many candies does Paul have now?
S1 O1 Q2	I'd like to know how many dishes there are now on the table (or "in the dresser"). There were ______ dishes on the table. Mammy layed down ______ dishes on the table. John layed down ______ dishes on the table.
S1 O2 Q1	On this morning ______ cars went into the garage. At noon ______ cars went into the garage. Yesterday there were already ______ cars in the garage. How many cars are there now in the garage?
S1 O2 Q2	I'd like to know how many sheets of paper does Aline have now in her exercise book. She has put in ______ pink sheets then she has put in ______ green sheets. She had already ______ white sheets in her exercise book.
S2 O1 Q1	Alan has now ______ marbles. During playtime, he won ______ marbles. At noon he won ______ marbles. How many marbles did Alan have before he went to school this morning?
S2 O1 Q2	I'd like to know how many persons there were in the playground before the coming of the boys and the girls. There are now ______ persons. ______ boys came, then ______ girls came.
S2 O2 Q1	At Christmas, Ann's grandmother bought ______ books for her. For the new year, her brother bought ______ books for her. Ann has now ______ books. How many books did Ann have before Christmas?
S2 O2 Q2	I'd like to know how many pencils did Valery have before receiving the red and the green ones. She received ______ red pencils then she received ______ green pencils. She has now ______ pencils.

cards were presented in random order). Finally, participants were given a short-term memory task with the number series of the Wechsler Intelligence Scale for Children.

When these warm-up exercises were completed, the experimenter told the child to try to solve the problems as quickly as possible, though without hurrying. The experimenter shuffled the cards on which the problems were written and picked one out. The problem was read, and the child was allowed to solve it. The experimenter then asked the child to give an answer, to explain how the problem was processed, and finally to recall the problem. Eight problems were presented following this procedure.

Results

Problems were scored according to the following schema: 0 = all wrong, 2 = all right, and 1 = right for a partially correct answer (one operation out of two). We adopted this scheme because of the hierarchical nature of the solution. To solve each problem, a middle operation had to be computed. Thus a participant could have stopped at this step and be credited with 1 point, or could have finished solving the problem and obtain a score of 2.

The analysis of variance (ANOVA) of these scores is equivalent to an ANOVA of binary data, and is known to be robust under these conditions (cf. Abdi, 1987; Hsu & Feldt, 1969).

The data were submitted to ANOVA with four fixed between-subjects factors. The four factors led to statistically reliable main effects. Contrary to our expectations, age led to increasing scores (6.25 at age 6, 7.28 at age 8, and 9.40 at age 10), $F(2, 168) = 9.52, p < .001$. Recall that, ideally, an age effect was not expected (due to the tentative adaptation of the problem difficulty to age). However, a finer grained analysis showed a strong interaction between age and structure, $F(2, 168) = 8.53, p < .001$ (see Figure 1).

As shown in Figure 1, this interaction is due essentially to the age factor having an effect only with S2 problems. An analysis conducted on two "subdesigns" confirms this interpretation: The weight of the age factor, negligible with S1, $\omega^2 = 0$, $F(2, 168) = 0.16$, ns, becomes very strong with S2, $\omega^2 = .97$, $F(2, 168) = 23.88$, $p < .00001$. Thus, equalizing the numerical series for the three age levels worked only for S1 problems. For S2, the results showed a clear-cut trend toward regular progress with age. These results can be explained by an increase in the abilities either of subtraction computation or of constructing an appropriate problem situation representation. Although our experiment does not allow us to decide between these alternatives, the current relevant literature supports the contention that the results cannot generally be attributed to an age improvement in subtraction ability; indeed, this operation does not appear to be more difficult than addition (Brainerd, 1983). Rather, the difficulties come from the obstacles created for the youngest children in constructing

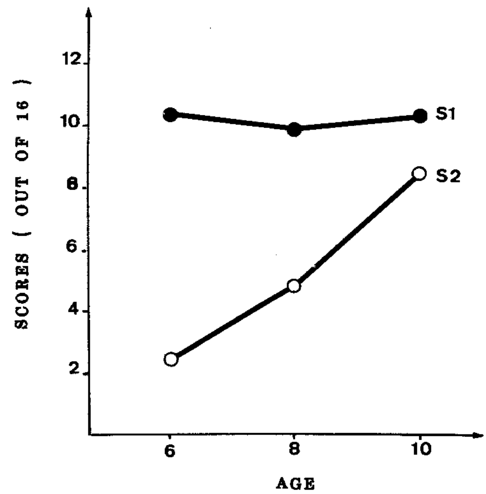

FIGURE 1 Age × Problem Structure interaction.

problem representations, a fact supported by previous research results (Riley et al., 1983; Vergnaud, 1982). Children's progress comes from a better understanding of problem-solving semantics.

For the problem types, S1 is clearly easier than S2 (*M*s = 10.26 and 5.09), $F(1, 168) = 79.58$, $p < .0001$. This is a commonplace result: In change problems, finding the result state always seemed easier than discovering the start state (Riley et al., 1983; Vergnaud, 1982). The order of information presentation proved to be significant, $F(1, 168) = 6.84$, $p < .01$. As anticipated, locating the transformations at the beginning of the problem text led to better performance than placing them after the state (*M*s = 8.44 and 6.91 for O2 and O1, respectively). The results supported our hypothesis; formulating the question first increased the scores, $F(1, 168) = 13.35$, $p < .001$ (*M*s = 8.73 and 6.62 for Q2 and Q1, respectively). Table 4 provides detailed results.

The Problem Type × Place of Question interaction was also statistically reliable, $F(1, 168) = 4.02$, $p < .05$. Although not anticipated, this interaction fits within our general theoretical framework. Thus, locating the question at the beginning of the problem text (Q2) led to better performance than locating the question at the end of the text (Q1); this difference was much more clear-cut under S2, when the problems are more difficult (see Figure 2 for detailed results).

EXPERIMENT 2[3]

Method

In this experiment, we used the same types of problems as those used in Experiment 1, although we simplified their numerical difficulty. Problems were made easier by using only two-digit numbers where the second digit was always 0 (as a consequence, there was no carrying). For example, if the numbers used in the first experiment were 24, 3, 5, the numbers in the second experiment were 20, 40, 10. We call the numbers of Experiment 1

[3]As the same control group is used in both studies, the statistical analysis of the experiment is not independent of the statistical analysis of the other study. A possible way to take into account this lack of independence is to treat both studies as two analytical comparisons (or multiple comparisons, or "subdesigns") derived from a complete experimental design. In this case, the use of the "Bonferroni approach" (cf. Abdi, 1987; Miller, 1981) gives a conservative statistical test. In our case, as we have two comparisons (or subdesigns), the Bonferroni approach is equivalent to dividing the significance level by 2 (i.e., to use alpha = .025 instead of .05 and alpha = .005 instead of .01). Although we have not reported this test, general interpretation remains unchanged when we adopt that procedure. Indeed, most of the reported significant results reached levels of significance smaller enough to be robust to any correction.

TABLE 4
Experiment 1: Mean Scores as a Function of Problem Type, Order of Information, Question Placement, and Age

	S1				*S2*			
	O1		*O2*		*O1*		*O2*	
Mean Age	*Q1*	*Q2*	*Q1*	*Q2*	*Q1*	*Q2*	*Q1*	*Q2*
6 years, 8 months	8.75	10.75	10.12	11.87	0.25	2.87	1.5	4.75
8 years, 8 months	8.5	9.5	11.5	10.37	0	6	5.37	7
10 years, 7 months	9.12	9.62	10.75	12.37	7.12	10.5	6.37	9.25

Note. Scores are of a possible 16.

"difficult numbers" (condition N2); the numbers of Experiment 2 are "easy numbers" (N1). We followed Experiment 1's procedures for Experiment 2.

We selected 64 new third graders (M = 8.4 years) with the same background as our first students to participate in this experiment.

Results

The data collected from the third graders were analyzed by ANOVA with four fixed between-subjects factors. Three of these factors proved statistically reliable. As in Experiment 1, we found strong main effects for problem type, $F(1, 112) = 75.8, p < .0001$, and place of transformations, $F(1, 112) = 28.94, p < .001$. The modifications affecting the place of the question did not show any reliable effect, $F(1, 112) = 1.08$. The novelty comes from the highly significant variations associated with simplifying the numerical series: mean score of 10.66 under N1 ("easy number") versus mean score of 7.27 under N2 ("difficult number"), $F(1, 112) = 25.1, p < .001$ (see Table 5 for detailed results).

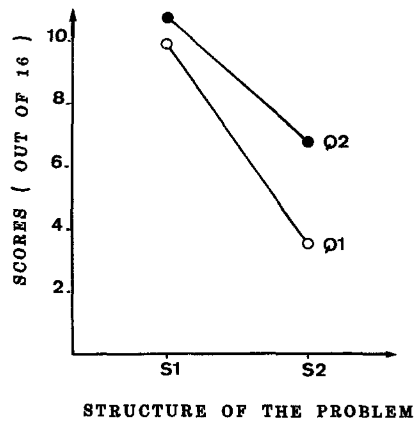

FIGURE 2 Problem Structure × Question Placement interaction.

TABLE 5
Experiment 2: Mean Scores as a Function of Number Series, Problem Type, Order of Information, and Question Placement

	N1 (Easy Numbers)				*N2 (Difficult Numbers)*			
	S1		*S2*		*S1*		*S2*	
Question Placement	*O1*	*O2*	*O1*	*O2*	*O1*	*O2*	*O1*	*O2*
Q1	13.87	14.12	3.37	12.25	8.5	11.5	0	5.37
Q2	13.25	14	2.75	11.5	9.5	10.37	6	7

Note. Scores are of a possible 16.

Three interactions reached significance. The first was only slightly reliable, $F(1, 112) = 3$, $p < .10$ (see Footnote 3), but, as anticipated, it showed that placing the question at the problem text's beginning led to better scores only with the difficult numbers (N2). This is consistent with our theory.

The other significant interactions were Problem Type × Order of Information, $F(1, 112) = 12.15$, $p < .001$, and the (just reliable) second-order Problem Type × Order of Information × Difficulty of Number Series, $F(1, 112) = 6.57$, $p < .05$. They showed that even if locating the two transformations at the problem text's beginning always led to better scores, this effect was systematically more prominent with the difficult problems (S2) associated with the "easy numbers" (N1). As can be seen in Figure 3, when all these conditions are combined (N1, S2, O2), the mean score almost reaches the one achieved under N1, S1, O2.

Nevertheless, we must interpret this last result cautiously. Indeed, as it

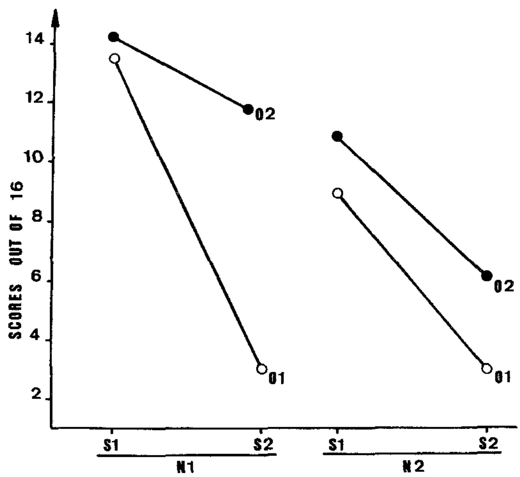

FIGURE 3 Second-order Number Series × Problem Structure × Order of Information interaction.

appears in Table 5, the mean score obtained under N1, S2, O1 is unusual in that it is the only case where a mean score ranks lower in N1 than in N2 ($M = 2.75$ under N1, S2, O1, Q1 vs. $M = 6$ under N2, S2, O1, Q1)—a fact that remains to be explained.

Thus, except for this last point, the quantitative analysis supported our hypothesis. An objection could be raised, however. In spite of the superior scores collected under N1, we cannot be sure that they were the result of better reasoning. They could have come from a more restricted number of errors in the computation under N1 in comparison with N2.

To rule out this objection, we conducted a qualitative analysis of the solution processes used by the third graders under N1 and N2. Due to the rare occurrences of semantic interpretation errors under S1, we studied only the S2 problems. The processes were inferred from the explanations given by the children and sometimes from their recollections.

We assessed and classified the processes according to their relevance to the problem's solution even if the numerical results associated with them were wrong. That is, we considered processes correct if they showed that the child had constructed a correct semantic interpretation. Concerning S2 problems, we identified three types of correct processes:

1. Compute $(T1 + T2) = T$ and then solve the subtraction $F - T$.
2. Compute $(F - T1) = M$ ("middle state") and then $(M = T2)$.
3. Compute $(T1 + T2) = T$ and then search for the cardinality of the complementary set $(T + x = F)$.

Although the proportion of correct processes decreased from N1 (51%) to N2 (42%), the amount of difference was smaller than expected. A fine-grained analysis shed some light on this fact. The observed performances were always better for N1 relative to N2 except for the combination O1, Q2—a point already mentioned and currently unexplainable. As Table

TABLE 6
Numbers and Proportions of Relevant Processes Used Under N1 and N2 Problems

	Difficulty of the Number Series	
Modalities of the Problems	*N1 (Easy Numbers)*	*N2 (Difficult Numbers)*
O1 Q1	13 (.20)	1 (.02)
O1 Q2	12 (.19)	35 (.55)
O2 Q1	56 (.88)	34 (.54)
O2 Q2	48 (.75)	37 (.58)

Note. Counts represent a possible 64 in each cell.

6 shows, a clear-cut improvement in the proportion of correct processes from N2 to N1 can be seen.

These facts lend support to our hypothesis. The qualitative analysis allowed us to separate, at least roughly, the difficulties due to computations from errors related to the semantic interpretation. It highlighted the anticipated interaction between the span of the numbers used in the problems and the level of reasoning: The harder the computations, the more numerous the errors in semantic interpretation. (However, Mann-Whitney tests failed to detect a significant difference between N1 and N2 under the same condition.) Indeed, we observed that when the numbers became difficult to store and process (as under N2) and when the wording of the text was not facilitative, a frequent error arose: S2 problems were treated as if they were S1 problems. This led to participants summing the three numerical data furnished, a fact we observed occurring at a rate of 84% under N2, O1, Q1 and at a rate of 34% under N1, O1, Q1.

GENERAL DISCUSSION

Consistent with our theory, we can see that the underlying semantic structure of verbally presented problem texts is one, but only one, factor involved in solving a problem. Indeed, as found elsewhere (Riley et al., 1983; Vergnaud, 1982), S2 problems (searching for an initial state while knowing final state and transformation) always appear more difficult than S1 problems. This proves true regardless of the age level or the difficulty of the numbers to be processed. We observed very large differences in our two experiments.

But the semantic structure is only one factor. As Kintsch and Greeno (1985) noted, "errors may reflect lack of knowledge, but at other times merely the limited information-processing capacity of the human organism" (p. 128). Children dealing with a verbally presented problem text must construct "on-line" a schematized problem representation. For that very reason, they depend on the text's wording.

Our results clearly show that slight variations in problem text organization lead to strong variations in scores, regardless of age. In particular, we found that placing two additive transformations and/or the question at the beginning of the problem text led to better performance. These findings were anticipated in light of our hypothesis, which stated that every factor lightening the WM load should facilitate the processing of the data at hand and thus, increase the performance level.

Interestingly enough, the effects of the textual organization manipulations were constrained by both ceiling and floor limits. The ceiling limit was encountered under S1 when the problems proved so easy that the children solved them whatever the conditions. The floor limit appeared in S2 with the first graders, when the problems seemed too difficult to be tackled in spite of any helpful information. Within these boundaries, we observed reliable effects as predicted by the theory.

We believe that placing the question at the problem text's beginning allows the participant to find out very quickly the relevant schematized problem representation and its associated problem-solving procedures. Thus, when such a representation is available, participants can proceed in a predominantly top-down fashion, filling in the empty slots of the schema with the data as they are encountered. In the Q1 condition, they are compelled to proceed in a more bottom-up manner, which overloads the WM space (see Kintsch & Greeno, 1985, for a processing model operating in a bottom-up fashion).

Placing the two additive transformations *T1* and *T2* at the beginning of the problem text also increases the scores. We believe this occurs because *T1* and *T2* are quickly and easily perceived as constituting a "chunk," and this is recorded under $T = T1 + T2$ after applying the addition. As T occupies less work-space than its two components, the WM load should be lightened. As a consequence, the problem's solution is facilitated.

The last factor to act in line with our theory is the relative difficulty of the numerical computations. As observed in Experiment 2, a decrement in difficulty (e.g., by substituting series like $20 + 40 + 10$ for $24 + 3 + 5$) to avoid carrying led to superior performance. That improvement does not stem wholly from a reduction of the computational error rate, but also from an increase in the proportion of relevant semantic interpretations. These results have been established with both a quantitative and qualitative analysis. However, confirmation by other experiments is still needed, insofar as some difficulties arose in a group of problems (N2, S2, O1, Q1). Nevertheless, most of the data confirm the theory's relevance: As the storage load increases in the WM system (in the "articulatory loop," following Hitch, 1980), the space devoted to reasoning (here constructing the schematized representation of the situation and applying the problem-solving procedures) is reduced. Consequently, as already observed in other tasks (Hitch, 1980), the error rate increases. Specifically, S2 problems tended to be solved as if they were S1 problems.

Thus it seems that solving verbally presented arithmetical problems requires that the student can either retrieve and activate in LTM or can construct in WM a relevant schematized representation of the problem situation. It also requires that the WM system, in which the numerical data are stored, rehearsed, and processed, is not overloaded.

This explains the effect of modifying the problem texts' wording: Lightening the load on WM increased the work-space devoted to reasoning, and better performance followed. Nevertheless, this is true only insofar as two conditions are fulfilled. First, the schemata of representation and solution must not be retrieved and applied in a fully automatic fashion. Indeed, as automatization rises, the WM becomes less and less taxed, a fact clearly observed with S1 problems. Consequently, even if work-space appears very limited, a solution remains possible. Second, the student must not be given a completely new or ill-mastered problem. If this happens, he or she can succeed but only insofar as he or she can mimic the actions or relationships described in the problem with material at his or her disposal (Carpenter, Hiebert, & Moser, 1981; Carpenter & Moser, 1982). All these procedures avoid overloading but cannot be used "in the mind." This is why we observed such a high error rate on the S2 problems, and with the third graders; WM was too heavily taxed to be able to "handle" the tasks of constructing a schematized representation of the problem situation.

To summarize, solving verbally presented arithmetical problems seems to require an adjustment between two kinds of processes. The first process functions mainly in a bottom-up manner; it is essentially a WM space-dependent process and so taxes the work-space. It could be fruitfully simulated by a computer model such as the one presented by Briars and Larkin (1984). This model mimics the problem-solving processes by acting them out with representation of physical counters. The second process acts in a top-down fashion providing ready-to-fill-in "schemata"; the data is essentially "knowledge-dependent." This process does not put an important load on WM capacity. Success depends on a relatively precise balance between these two processes. Finally, even if work-space appears very limited, its storage and processing capacities depend on knowledge and practice as much as on a fixed dimension (developing or not).

ACKNOWLEDGMENTS

We thank L. B. Resnick, J. G. Greeno, A. J. O'Toole, and two anonymous reviewers for their helpful suggestions and for their comments on previous versions of this article.

REFERENCES

Abdi, H. (1987). *Introduction au traitement statistique des données experimentales* [Introduction to experimental data analysis]. Grenoble, France: Presses Universitaires de Grenoble.

Ashcraft, M. H. (1982). The development of mental arithmetic: A chronometric approach. *Developmental Psychology, 2,* 213–236.

Ashcraft, M. H., & Battaglia, J. (1978). Cognitive arithmetic: Evidence for retrieval and decision processes in mental addition. *Journal of Experimental Psychology: Human Learning and Memory, 4,* 527–538.

Ashcraft, M. H., & Fierman, B. A. (1982). Mental addition in third, fourth, and sixth graders.

Journal of Experimental Psychology, 33, 216–234.

Baddeley, A. D. (1986). *Working memory.* Oxford: Oxford University Press.

Baddeley, A. D., & Hitch, G. (1974). Working memory. In G. H. Bower (Ed.), *The psychology of learning and motivation* (pp. 47–88). New York: Academic.

Brainerd, C. J. (1983). Young children's mental arithmetic errors: A working memory analysis. *Child Development, 55,* 812–830.

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). Problem structure and first grade children's initial solution process for simple addition and subtraction problems. *Journal for Research in Mathematical Education, 12,* 27–39.

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

Case, R. (1982). General development influences on the acquisition of elementary concepts and algorithms in arithmetic. In T. P. Carpenter, J. M. Moser, & T. A. Romberg (Eds.), *Addition and subtraction: A cognitive perspective* (pp. 157–170). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Case, R., Kurland, D. M., & Goldberg, J. (1982). Operational efficiency and the growth of short-term memory span. *Journal of Experimental Child Psychology, 33,* 386–404.

Dempster, F. N. (1981). Memory span: Sources of individual and developmental differences. *Psychological Bulletin, 99,* 63–100.

Groen, G. J., & Parkman, J. M. (1972). A chronometric analysis of simple addition. *Psychological Review, 79,* 329–343.

Hitch, G. J. (1978). The role of short-term working memory in mental arithmetic. *Cognitive Psychology, 10,* 302–323.

Hitch, G. J. (1980). Developing the concept of working memory. In G. Claxton (Ed.), *Cognitive psychology: New directions* (pp. 154–196). London: Routledge & Kegan Paul.

Hitch, G. J., & Halliday, M. S. (1983). Working memory in children. *Philosophical Transactions of the Royal Society of London, 8,* 325–340.

Hsu, T., & Feldt, L. T. (1969). The effect of limitations on the number of criterion score values on the significance level of the *F*-test. *American Educational Research Journal, 6,* 515–527.

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

Hulme, C., Thomson, N., Muir, C., & Lawrence, A. (1984). Speech rate and the development of short-term memory span. *Journal of Experimental Child Psychology, 38,* 241–253.

Kieras, D. E. (1980). Initial mention as a signal to thematic content in technical passages. *Memory and Cognition, 8,* 345–353.

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

Kosminsky, E. (1977). Altering comprehension: The effect of biasing titles on test comprehension. *Memory and Cognition, 5,* 448–482.

Miller, R. G. (1981). *Simultaneous statistical inferences* (2nd ed.). Berlin: Springer-Verlag.

Naus, M. J. (1982). Memory development in the young reader. The combined effects of knowledge base and memory processing. In O. Wayne & S. White (Eds.), *Reading expository material* (pp. 49–74). New York: Academic.

Pascual-Leone, J. (1970). A mathematical model for the transition rule in Piaget's developmental stages. *Acta Psychologica, 32,* 301–345.

Riley, M. S., Greeno, J. G., & Heller, J. I. (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.

Svenson, D. (1975). Analysis of time required by children for simple additions. *Acta Psychologica, 39,* 289–302.

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. A. Romberg (Eds.), *Addition and subtraction: A cognitive perspective* (pp. 39–59). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.