Working memory and intrusions
of irrelevant information in a group
# of specific poor problem solvers

MARIA CHIARA PASSOLUNGHI
University of Trieste, Trieste, Italy

and

CESARE CORNOLDI and STEPHANIE DE LIBERTO
University of Padua, Padua, Italy

An important body of evidence has shown that reading comprehension ability is related to working memory and, in particular, to the success in Daneman and Carpenter's (1980) reading and listening span test. This research tested a similar hypothesis for arithmetic word problems, since, in order to maintain and process the information, they require working memory processes. A group of children possessing average vocabulary but poor arithmetic problem-solving skills was compared with a group of good problem solvers, matched for vocabulary, age, and socioeconomic status. Poor problem solvers presented lower recall and a greater number of intrusion errors in a series of tasks testing working memory and memory for problems. The results obtained over a series of six experimental phases, conducted during a 2-school-year period, offer evidence in favor of the hypotheses that groups of poor problem solvers may have poor performance in a working memory test requiring inhibition of irrelevant information (Hypothesis 1), but not in other short-term memory tests (Hypothesis 2), that this difficulty is related to poor recall of critical information and greater recall of to-be-inhibited information (Hypothesis 3), that poor problem solvers also have difficulty in remembering only relevant information included in arithmetic word problems (Hypothesis 4) despite the fact that they are able to identify relevant information (Hypothesis 5). The results show that problem-solving ability is related to the ability of reducing the memory accessibility of nontarget and irrelevant information.

The present study examined the relationship between working memory and arithmetic word problem solving. This issue reflects an increasing interest in the role of cognitive abilities in problem solving (for reviews, see Lucangeli & Passolunghi, 1995, and Mayer, 1992). In the last decade, a variety of theoretical frameworks has been proposed to describe the cognitive processes necessary to solve a word problem. Many of these theories focused on the comprehension and construction of a representation of the problem (e.g., Nathan, Kintsch, & Young, 1992; Passolunghi, Lonciari, & Cornoldi, 1995) and of control processes (e.g., Lucangeli & Cornoldi, 1997; Pressley,

1990): All these processes seem to involve the use of working memory.

These results are consistent with further evidence suggesting that working memory can be involved in different aspects of arithmetic (Leather & Henry, 1994; Logie, Gilhooly, & Wynn, 1994; Passolunghi, Cornoldi, & De Liberto, 1996; Siegel & Ryan, 1989; Swanson, Cooney, & Brock, 1993). Indeed, the different operations required in arithmetic tasks typically require that the subject maintain some information in short-term memory (e.g., data, task requests, operations, partial results, etc).

The controlled components of working memory can be involved in arithmetic word problem solving even when the written text is still available. Comprehension of a text implies that the incoming information is integrated with the previous information maintained in the working memory system (Baddeley, 1990; Cornoldi, De Beni, & Pazzaglia, 1996; Oakhill & Yuill, 1996). Furthermore, the complete comprehension of the problem requires that the solvers build up a mental representation that involves the capacity of the working memory system. From the point of view of Baddeley's (1986) three-components model, the central executive could be more specifically and strongly involved than the articulatory loop. Indeed, problem solving does not simply involve the maintenance of given in-

formation but requires its control—that is, that this information is examined for relevance, selected or inhibited (according to its relevance), integrated, used, and so on (see also Turner & Engle, 1989).

The assumption of a relationship between working memory and children's word problem solving has been considered by Swanson and his coauthors in a series of studies. In particular, Swanson et al. (1993) tested third and fourth graders with normal achievement in a series of memory tasks. As a working memory measure, they used a modification of the Daneman and Carpenter's (1980) reading span task. This task requires the child to read and semantically process an increasing set of sentences and to memorize the last word of each sentence. Typical measures considered for the reading span test are accuracy in answering the verification questions, number of sentence-final words correctly recalled, and the span level (i.e., the highest number of sentence-final words correctly recalled). Swanson et al., using the proportion of words retrieved in each set as a measure of working memory, found that working memory ability was significantly correlated with problem-solving accuracy; however, the relationship between the two measures was low, especially when the influence of other abilities was partialed out. In a further study, Swanson (1994) proposed the listening span task, which has the same procedure as the reading span task described above with the only difference being that the subject listens to the sentences and does not read them.

An increasing body of evidence shows that inhibition can play a role in reading comprehension. De Beni, Palladino, Pazzaglia, and Cornoldi (1998) pointed out that the relationship between listening/reading span and reading comprehension could be attributed to a failure of inhibition, identified by an inability to control intrusive, irrelevant information. Similarly, Hasher and colleagues (e.g., Hasher, Stoltzfus, Zacks, & Rypma, 1991; Hasher & Zacks, 1988) suggested that elderly people have poor inhibition and have difficulty in selecting and retrieving the antecedent information necessary to form an inference. With a different line of research, Gernsbacher and colleagues (e.g., Gernsbacher, 1993; Gernsbacher & Faust, 1991) found that less skilled readers have less efficient suppression mechanisms. For example, in a task presenting sentences with ambiguous words, less skilled comprehenders are slower to reject the inappropriate meaning of ambiguous words (see also Meiran, 1996). Altogether, these results suggest that a difference in working memory span might be related not to the quantity of information that working memory can hold at the same time but to the efficiency of an inhibition mechanism that eliminates irrelevant or no-longer-relevant information from working memory.

The aim of this study was to extend this line of research to the relationship between working memory and problem solving, suggested also by the research on the role of control processes (Montague, 1992; Pressley, 1990) and selection of relevant information in problem solving (Low

& Over, 1992; Low, Over, Doolan, & Mitchell, 1994). However, this indirect evidence requires more direct proof of a relationship between the ability to solve arithmetic word problems and the ability to control irrelevant information in working memory. In our view, word problem solving requires that working memory hold the relevant information conveyed by the problem. In order to do this effectively, the person has to reduce the accessibility of less relevant information, which could overload working memory and interfere during the processing of the information. Subjects with poor problem-solving ability could have difficulty controlling irrelevant information.

In the present study, we compared the memory performance of groups of children who were poor at problem solving and those who were good at problem solving. This methodology has been used in a few studies, in particular by Siegel and colleagues (Shafrir & Siegel, 1994; Siegel & Linder, 1984; Siegel & Ryan, 1988), who found that children with specific difficulties in arithmetic presented memory difficulties and, in particular, relied to a lesser extent on short-term memory phonetic encoding than did control groups.

We examined this hypothesis in a selected group of fourth graders, poor and good at arithmetic word problems, tested with a series of working memory tasks, over a 2-school-year period. The group of poor problem solvers met criteria that made it possible to reject the risk that possible differences found between them and a comparison group of good problem solvers were due to other concurrent factors rather than working memory.

In the first two phases of our study, we verified that the group of poor problem solvers performed worse than did the group of good problem solvers in listening spans. We hypothesized that this difference was caused by the children having difficulty in avoiding memory intrusions of irrelevant information (i.e., the parts of the sentences, or nouns, which are not to be recalled). In a third phase, we hypothesized that we might find the same pattern of data in a task requiring the subjects to remember the information contained in arithmetic word problems. In a fourth phase, we checked the subjects' capacity to distinguish the relevant and irrelevant information included in the text of the word problems and their working memory/short-term memory ability. In the fifth and sixth follow-up phases, we ruled out the possibility that some maturation effect (concerning memory and mathematical abilities), over the course of the two school years, might have affected the results. Moreover, in the sixth phase, we also looked at the recognition test of irrelevant information presented in a working memory test similar to that used in the second phase.

THE LONGITUDINAL STUDY

The subjects were 33 fourth graders divided into a group of 15 poor problem solvers and a group of 18 good problem solvers. These groups were formed on the basis of a wide-scale screening proposed to a group of 300 stu-

Table 1
Mean Performance and Standard Deviations of the Group of Poor Problem Solvers and the Group of Good Problem Solvers on a Working Memory Listening Span Test

	Poor Problem Solvers		Good Problem Solvers	
Index	M	SD	M	SD
T/F	17.67	0.62	17.89	0.32
Correct Rec.	10.47	2.36	13.22	2.13
Intrus.	3.60	2.50	1.11	1.02
Sequen.	0.53	0.74	1.72	1.45

Note—T/F, number of correct true/false responses in the verification of sentences; Correct Rec., number of correctly recalled sentence-final words; Intrus., number of intrusions (i.e., words recalled despite the fact that they were not at the end of a sentence); Sequen., number of sequences of sentence-final words that were completely recalled.

dents from two different cities in Northern Italy (Milano and Trieste), during the fall semester of their fourth grade. In this screening, we collected, in group testing administered in the classes, scores on arithmetic word problem-solving achievement in a standardized test (Amoretti, Bazzini, Pesci, & Reggiani, 1994), the vocabulary subtest of the PMA battery (Thurstone & Thurstone, 1941), which is considered a measure of verbal intelligence, and we asked teachers to indicate the children's problems related to behavior, arithmetic problem solving, emotionality, and sociocultural level. The standardized mathematical test included 12 short word written arithmetical problems. One example is the following: "On Pascoli Street there are 45 shops. ⅗ of them sell clothes. How many clothes shops are there in Pascoli Street?" Children were included in the group of poor problem solvers if they were in the 20th percentile or below on the mathematical test and the teachers had mentioned corresponding problems, whereas the good problem solvers were between the 50th and the 80th percentile on the same mathematical test and were considered to be without any particular problem in mathematics by their teachers. The two groups were matched for sociocultural level, behavior, and their scores on the vocabulary test. Mean scores on the vocabulary test were respectively 51.5 (SD = 5.7) for the group of poor problem solvers and 54.2 (SD = 5.3) for the group of good problem solvers, the difference not being statistically significant. In particular, the IQ of children, as inferred from the vocabulary score, was never lower than 90. The mean age of the two groups was the same (9 years and 6 months).

Children selected in this screening were tested in three sessions between January and April of their fourth grade. At the beginning of their fifth grade, between late September and early October, the children underwent a fourth test session. Finally, the same children (except a few children who could not be retested) were tested a fifth and a sixth time during the last 2 months of their fifth grade.

Phase 1

Phase 1 tested the hypothesis that poor problem solvers have a worse listening span than children good at arithmetic word problem solving. Furthermore, it tested the hypothesis that poor problem solvers were less able to eliminate the not-to-be-remembered items (i.e., words that were not placed at the end of the sentence).

Method

Materials. We devised an Italian adaptation of the listening span test of Daneman and Carpenter (1980). Stimuli were reduced in number and complexity in order to be age-appropriate for young children. A pilot administration of the procedure to groups of children of the same age confirmed that the test was easily understood and was able to differentiate children. The materials consisted of two series of two sentences, two series of three sentences, and two series of four sentences. For each two-series set, there was a practice series. The sentences were either true statements or false statements. An example of true sentence was "The cow has four legs and one TAIL"; an example of false sentence was "Eyeglasses help to better listen to the SOUNDS."

Procedure. The experimenter presented the task and checked the child's comprehension. Each sentence was presented at a rate of approximately 1 sec per content word, followed by an interval for the child's answer concerning the truth of the sentence. The child was instructed to give the answer as soon as possible. His/her answer was followed by the presentation of the next sentence. At the end of each set of sentences, immediately after the verification answer for the last sentence, the child was asked to recall the last word of each sentence. The subjects were instructed to avoid mentioning non–sentence-final words. For example, the correct responses were "TAIL" and "SOUNDS" for the sentences mentioned above.

Results

The mean performance on the listening span task of the two groups is presented in Table 1. Both groups of children did not have particular difficulty in the verification task (the maximum possible score was 18). All the children had difficulty in remembering the complete sets of final words. The good problem solvers had a mean span score (i.e., the number of sequences of sentence-final words that were completely recalled) that was three times greater than that of the poor problem solvers (1.72 vs. 0.53). The difference was significant, as shown by a t-test comparison for unequal variances [$t(26.27) = 3.04, p = .005$] (variances were deemed significantly different by the Levene's test). Good problem solvers recalled more nouns correctly [$t(31) = 3.53, p = .001$] and more nouns when order was considered. The mean number of order errors was similar in the two groups (3.87 for the group of poor problem solvers and 3.39 for the group of good problem solvers). Despite the fact that the poor problem solvers recalled fewer correct nouns, they erroneously remembered three times more non–sentence-final words than the good problem solvers [$t(31) = 3.80, p < .001$].

Since information on reading comprehension ability (scores in a standardized test; Cornoldi & Colpo, 1981) was made available at the end of the school year for 14 poor problem solvers and 17 good problem solvers, we examined its implications for our results. The good problem solvers had significantly better scores ($M = 9.00, SD = 1.37$) than the poor problem solvers ($M = 7.86, SD = 1.35$) in the standardized reading comprehension test [$t(29) = 2.33, p = .027$]. However, when we used comprehension

as a covariate of the listening span test in an analysis of covariance, we still found a significant difference between the groups, both for the number of correct responses [$F(1,28) = 16.18, p < .001$] and for the number of intrusions [$F(1,28) = 13.05, p = .001$].

Discussion

The results of this phase showed that the children with low arithmetic word problem-solving abilities had a lower performance in the listening span test. A consistent body of evidence had shown that reading comprehension ability is related to success in the listening span test. Evidence related to problem-solving ability was still required, since the only available data, concerning a group of normal youngsters tested by Swanson et al. (1993), did not completely clarify the problem. The more clear-cut data found in this research might have been due to differences either in the groups or in scoring and procedure.

Successful performance on the listening span task requires that the subject avoid intrusions in memory of non–sentence-final words. Good problem solvers were better able to avoid intrusion errors. It is worthwhile to note that, if we sum the number of correct last words and the number of intrusions, the total number of words recalled was similar for the poor problem solvers (14.07) and the good ones (14.33). Therefore, we hypothesize that the differences between the groups' performances could have been due to an inability to inhibit the non–sentence-final words.

The poor problem solvers' performance in the listening span might be attributed to reasons other than inhibition difficulties. For example, it could be argued that the poor problem solvers had a memory problem (e.g., they forgot the information order and, therefore, were not able to discriminate between items placed in the last position in the sentences and items in other positions). However, the slight difference in the number of order reversals errors (this number was even lower in the case of poor problem solvers) suggests that performance differences were not due to a greater loss of order information in the group of poor problem solvers.

Another possibility for why poor problem solvers scored lower on the listening span test is that they may have had more difficulty in processing sentences. A similar semantic-elaboration hypothesis was proposed to explain the relationship between reading comprehension abilities and listening span (Daneman & Carpenter, 1980, 1983), which could be extended to the case of children weak in arithmetic word problem solving. Indeed, language comprehension is related to arithmetic word problem solving, and children weak in arithmetic word problem solving might also be weak in language comprehension. Despite the fact they were able to verify the proposed sentences, the poor problem solvers might have had more difficulty in the task, thus drawing more resources from the general working memory system and then leaving less resources for the span task. In the second phase, we

checked this aspect. Since our groups were matched for a vocabulary test involving lexical competence, in Phase 2, we thought it was appropriate to use a memory task, using single words instead of sentences, devised for examining poor comprehenders (De Beni et al., 1998).

Phase 2

We presented a revised listening span test to the two groups that required the subjects to remember the last item of different series of words. The subjects listened to series of strings of words (rather than series of sentences) and performed a simple categorization task (i.e., they tapped the table when the name of an animal was presented). Only concrete highly familiar words (including names of animals) were presented.

Memory performance is affected by the responses required in the orienting task. For example, Craik and Tulving (1975) found that items subjected to a deeper semantic encoding on the basis of the request of a semantic judgment had higher memory activation if the semantic judgment was affirmative. Therefore, we hypothesized that animal nouns might be more difficult to inhibit and increase the number of intrusions in poor problem solvers, as has been shown for poor readers (De Beni et al., 1998).

Method

Materials. The materials consisted of 12 sets of stimuli, each composed of a string of four words. The strings were used in place of the sentences typically used in the listening span test. There were 4 sets with 2 strings, 4 sets with 3 strings, and 3 sets with 4 strings, for a total of 32 strings and 128 words. Of these words, 36 were animal nouns (10 in the final positions, and 26 within the string). An example of a string was the following: *stone, desk, father, mouse.*

Procedure. The children were individually tested a few days after the first testing of Phase 1, following a similar procedure. The words were presented at an approximate rate of 1 sec per word, with a 2-sec interval after each string of words (the end of each string was indicated with a stress in the voice, after the last word in the string). When the name of an animal was given, the child had to tap the table. At the end of each series, the child was asked to recall the last word in each string in the series.

Results

A small group of children (4 subjects in the group of poor problem solvers and 4 subjects in the group of good problem solvers) could not be tested in the entire sequence, and, thus, 3 series had to be dropped. For this reason, we computed the overall performances in mean proportion. Table 2 presents the mean proportion of sequences of the last items that were completely recalled in the right order, the mean proportion of the correct last items, the mean proportion of animal intrusions, and the mean proportion of other intrusions.

Poor problem solvers remembered fewer completely correct sequences [$t(31) = 2.56, p = .016$] and fewer words in the last position [$t(31) = 2.24, p = .032$]. Further analysis counted the number of animal names situated in the last position that were correctly recalled: animal names were remembered very well (75% correct recall [$SD = $

Table 2
Mean Proportions and Standard Deviations of
Different Items Recalled in the Word Listening Span Test
by the Poor Problem Solvers and the Good Problem Solvers

Index	Poor Problem Solvers		Good Problem Solvers	
	M	SD	M	SD
Sequen.	.10	.08	.25	.24
Correct Rec.	.44	.11	.56	.13
Animal Intrus.	.13	.09	.06	.09
Other Intrus.	.09	.08	.03	.04

Note—Sequen., proportion of sequences that were completely recalled; Correct Rec., proportion of correctly recalled list-final words; Animal Intrus., proportion of animal (name) intrusions (i.e., words recalled that were not at the end of the strings); Other Intrus., proportion of nonanimal intrusions (i.e., recalled words that were not animal names and were not at the end of the strings).

18] by good problem solvers and 72% correct recall [SD = 10] by poor problem solvers); there was significantly more correct recall of animal names than of other words (p = .001). The difference in the recall of animal names between the two groups was not significant (p = .50). In contrast, the poor problem solvers recalled fewer nonanimal words situated in the last positions [$t(31)$ = 2.95, p = .003].

The comparison of correct recall of animal names in the last position versus the recall of nonanimal names in the last position (the results of the groups of poor and good problem solvers taken together) clearly indicates that animal names were particularly well remembered: 74% correct recall (SD = 15) of animal names in the last positions and 62% correct recall (SD = 20) of nonanimal names in the last position; the difference was significant [$t(32)$ = 5.005, p = .0001].

The results concerning the intrusions presented the same data pattern as in Phase 1. Poor problem solvers made more intrusion errors. Furthermore, intrusion errors were particularly prevalent with animal names. A 2 × 2 analysis of variance (ANOVA) for a mixed design on the number of intrusions revealed a significant main effect of groups [$F(1,31)$ = 7.82, MS_e = 0.06, p = .009] and a significant main effect of intrusion types [$F(1,31)$ = 4.85, MS_e = 0.02, p = .035], due to more errors with animal names than with other nouns.

Discussion

The results of Phase 2 extend the generality of the listening span effect found in Phase 1. Poor problem solvers scored lower than good problem solvers on the listening span task even when the role of sentence comprehension was eliminated. The verification task was particularly easy for both groups; they did it without any difficulty and with no mistakes. The span task was perfectly understood by all subjects but created some difficulties. In particular, poor problem solvers had difficulty in remembering the last words of the word strings, reporting less than half of the correct stimuli and reporting a significant number of nontarget items.

This phase offers more evidence against the hypothesis that intrusion errors are related to memory degradation: Animal names had higher memory strength than the other names but, nevertheless, produced more intrusions. It is interesting to note that the two groups of problem solvers were similar in the recall of animal names, whereas the poor problem solvers had lower overall performance with nonanimal names.

With respect to memory intrusions, both groups produced more animal names as intrusion errors than nonanimal words. This suggests that inhibition is particularly difficult for items subjected to more extended encoding (Craik & Tulving, 1975).

Phase 3

Phase 3 tested the conclusions that poor problem solvers are characterized by inadequate inhibitory processing in a more direct way (i.e., looking directly at the subjects' memory for more relevant and less relevant information in arithmetic word problems). Memory for problems has not been extensively studied, but the literature on the subject offers some evidence in favor of the hypothesis that memory can vary in selectivity (Hegarty, Mayer, & Monk, 1995; Mayer, 1982).

We predicted that the poor problem solvers would have worse recall of relevant information and greater recall of irrelevant information than the good problem solvers.

Method

Materials. We devised two parallel series of 12 problems. The problems were taken from typical arithmetic word problem sets in fourth-grade classes. An example of a problem was the following:

Four good friends go to a "pizzeria." Each of them eats a pizza which costs 8,500 Liras and orders a drink which costs 2,500 Liras. What does the bill come to? If one of them pays with a bill of 50,000 Liras, how much change will he receive?

Procedure. The children were individually tested in two separate sessions, the second approximately 1 week after the first. In each session, they were presented with 6 of the 12 problems from one series. One of four possible presentation order problems was randomly assigned to each subject. The experimenter informed the child that he/she had to listen carefully to each problem in order to remember the problem, focusing on the relevant information included in the problem. The instructions included a short example and questions in order to check that the child had understood the task. The problem was read to the child once at the standard rate using a normal intonation (e.g., the problem presented above took about 35 sec). At the end of each problem presentation, the child was asked to recall its relevant information, and the child's answers were tape recorded. At the end of this task, the child was given a booklet with the 6 problems presented in that session and was asked to solve them. The children were allowed as much time as needed to solve the problems.

Results

The subjects' recall protocols were scored by calculation of the relevant and irrelevant information recalled. For example, the "good friends" and the "pizzeria" mentioned in the "pizza" problem were considered irrelevant, whereas the prices of the pizza and drink, the number of

Table 3
Mean Scores and Standard Deviations of the
Poor Problem Solvers and the Good Problem Solvers in
the Memorization and Solution of Arithmetic Word Problems

Index	Poor Problem Solvers		Good Problem Solvers	
	M	SD	M	SD
Relevant	30.13	9.94	39.94	6.68
Irrelevant	9.00	6.22	5.17	3.75
Correct Probs.	2.53	2.64	8.78	2.24
Correct Ops.	7.87	3.80	23.67	3.80

Note—Relevant, number of relevant elements recalled; Irrelevant, number of irrelevant elements recalled; Correct Probs., number of problems correctly solved; Correct Ops., number of arithmetic operations correctly indicated.

friends, the value of the bank note, and the question were considered relevant. Table 3 presents the mean scores of the relevant and irrelevant information recalled by the two groups. A 2 × 2 ANOVA for a mixed design, groups × kind of information (relevant, irrelevant), revealed a significant group effect $[F(1,31) = 4.574, MS_e = 31.96, p = .04]$, a significant effect of information kind $[F(1,31) = 204.94, MS_e = 62.40, p = .0001]$, and a significant interaction $[F(1,31) = 12.205, MS_e = 62.40, p = .0015]$.

The poor problem solvers recalled less relevant information and more irrelevant information. Both differences were significant [relevant information, $t(31) = 3.37, p = .002$; irrelevant information, $t(31) = -2.19, p = .03$].

The data on problem solving (Table 3) confirmed that the poor problem solvers had difficulty in solving problems. They solved fewer problems perfectly than did the good problem solvers $[t(31) = 7.35, p < .001]$ and made fewer correct choices of the operations required by the problems $[t(21.40) = 8.17, p < .001]$. Analysis of the subjects' protocols found that the correctly solved problems corresponded frequently to the problems that were more adequately remembered. More precisely, we observed that recall for correctly solved problems was approximately double that for unsolved ones.

Discussion

The results of Phase 3 show that problem-solving ability is related to the capacity to remember the relevant information contained in a problem. Whereas a subject usually has the written text of a problem available, we note that, for our memory task, the problem was presented auditorily and was given in a written version to the subject only when he/she had to solve it. Despite this difference, problem memory was related to problem-solving ability. This result suggests that the working memory operations required in this task are also critical in problem solving. In particular, having to base their memory on a single auditory presentation, the subjects could not remember the whole text of the problem and were forced to focus on its most relevant elements and to eliminate the less relevant elements. These results are congruent with Hegarty et al.'s (1995) research showing that successful problem solvers

are more efficient and selective in problem recall. On the contrary, the data are not congruent with Swanson et al.'s (1993) observation that problem-solving ability is not related to memory differentiation between relevant and extraneous information. However, it is possible that Swanson et al.'s task motivated the subjects to study the overall text, adopting a strategy capable of memorizing all the information. This strategy is useful in study tasks, but it is not critically related to the problem-solving process.

Our task largely mirrored the tasks we used in the preceding phases. The subjects had to listen to auditorily presented information and to recall relevant information. We can assume that the subjects processed all the material: Comprehension of the problem requires that all the sentences are to be processed (Mayer, Larkin, & Kadane, 1984). Therefore, as in the span tasks used in Phases 1 and 2, the subjects had to select target (relevant) information and to inhibit (irrelevant) information, which they had earlier processed. However, unlike the preceding span tasks, where target and nontarget information was clearly defined on the basis of a simple rule, in this task, the subjects had to determine the relevant information themselves. Therefore, our results might have been due to the poor problem solvers being unable to identify the irrelevant information included in the problems, rather than being unable to reduce its accessibility. In order to examine this issue, we ran a further phase.

Phase 4

Phase 4 tested our subjects' capacity to differentiate between the relevant and irrelevant information in arithmetic word problems. For each of a series of problems, presented in written format, the children had to underline the parts they thought were most relevant. Furthermore, they were asked to write down the problem data, a procedure taught in Italian schools.

A second check concerned our children's working memory abilities. Our data showed that the poor problem solvers had, in general, good memory, having (on the whole) as much recall as the good problem solvers. However, some researchers have shown that disabled child learners have a deficit in short-term memory, and this also applies to children with specific mathematical problems (e.g., Siegel & Linder, 1984; Siegel & Ryan, 1988).

In Phase 4, we examined our children's memory by using two versions of a word span test: a simple version and a dual-task version. Span tests are common means of examining short-term memory, and their psychometric properties have been largely demonstrated with slightly different procedures and materials (e.g., Reynolds & Biegler, 1987). We used familiar words rather than digits (used in some experiments with children with specific difficulty with arithmetic calculation; e.g., Hitch & McAuley, 1991) in order to have a task directly comparable with the task used in Phase 2. The span test was implemented in a simple standard version (simple auditory

Table 4
Mean Proportions and Standard Deviations of Relevant and Irrelevant Problem Information Indicated by the Poor Problem Solvers and the Good Problem Solvers, in an Underlining and in a Data Selection Task, and the Mean Performances of Both Groups in Two Span Tests

Index	Poor Problem Solvers		Good Problem Solvers	
	M	SD	M	SD
Underlining				
Relevant	.95	.10	.99	.02
Irrelevant	.31	.30	.21	.23
Data				
Relevant	.96	.08	.98	.04
Irrelevant	.16	.23	.6	.07
Span Tests				
Simple Span	4.13	0.52	4.12	0.33
Dual Span	4.33	0.62	4.18	0.39

presentation and immediate repetition of the words) and in a dual-task version. In the dual-task version, the child was required to tap the table when the name of an animal was presented (during the presentation of the stimuli). This task exactly replicated the task used in Phase 2. The two versions provided us the possibility to test whether the results obtained in Phase 2 were due to poor problem solvers' difficulty in paying attention to two simultaneous requests.

In Phase 4, we were able to test the same children who had participated in the previous three phases, with the exception of 1 of the good problem solvers, who had moved to another school. The subjects were tested while they were attending the first months of their fifth grade.

Method

Materials. The materials were six new arithmetic word problems, similar to the problems used in the preceding phase and corresponding to the school level reached by the subjects, including both relevant and irrelevant information. An example of a problem was the following:

> During the summer Marisa helped her parents at home by doing the cleaning and as a reward received 22,500 Liras. Marisa was very greedy and decided to buy 12 special almond sweets. Each sweet costs 1,500 Liras. How many almond sweets can Marisa afford to buy with the remaining money?

For the memory task, we used different strings of familiar words, varying in length from three to eight words per string (two strings for each length) for a simple word span task and a similar set of strings for the dual word span task.

Procedure. The children were tested individually. The presentation order of the tasks was balanced between subjects, in such a way that approximately half of the subjects started with the problems and then did the simple span task, followed by the dual span task, whereas the other half started with the simple span task and then did the dual span task and, finally, the problems. In the problems situation, the children received a booklet with one problem per page. They were required to underline the relevant information contained in the problem (i.e., the elements necessary to solve the problem). The children also wrote below the text the data for each problem as they were taught by their teacher. A typical example of data description required in Italian schools is the following: (for the example of "Marisa's" problem, see the Materials section above)

22,500 Lira = Marisa's money; 12 = number of sweets; 1,500 Lira = cost of each sweet.

In the simple span task, strings of words of increasing length were presented at a rate of 1 sec per word until the child made a mistake for both series of the same length. A string was considered to have been recalled if all the nouns were recalled in their correct order. The experimenter started with a string of three nouns. If the repetition was correct, the experimenter moved to a longer string; if the repetition was incorrect, the child had the opportunity of trying again with another string of the same length. In the dual span task, the procedure was identical, but, in each string, there was an animal word, and, in correspondence with its presentation, the child had to tap the table. The animal word could be in any position except the last one, in order to avoid the tapping delaying the repetition of the words.

Results

Table 4 presents the mean performance of the two groups involved in the tasks. The two groups performed similarly on the span tasks (in fact, the poor problem solvers were even slightly better), with no significant difference between the two groups. The standard span test was always presented before the dual task because we were concerned that the children would confuse the instructions. For this reason, direct comparison between the two span tests is potentially confounded by order effects. Nevertheless, the data suggest that the dual task did not create any interference for either good problem solvers or poor problem solvers, since the performance with the dual span was no lower than the performance with the simple span.

In the underlining task, both groups were able to differentiate between relevant and irrelevant information. An inspection of the data (Table 4) makes this point self-evident. A 2 × 2 ANOVA (groups × relevance) revealed a significant relevance effect [$F(1,30) = 226.13$, $MS_e = 0.04$, $p < .001$]. No significant group ($p = .62$) or interaction ($p = .15$) effects were found. Similarly, for the data selection task, there was a significant relevance effect [$F(1,30) = 792.85$, $MS_e = 0.01$, $p < .001$], there was no group difference effect ($p = .25$), but there was a marginal interaction effect [$F(1,30) = 3.52$, $p = .07$].

Discussion

The data on the span tests are compelling, since the poor problem solvers were not significantly worse than the good problem solvers but were even slightly better. This confirms that our poor problem solvers' short-term memory ability was as good as that of the good problem solvers. Furthermore, the data using the dual span task show that the tapping request, introduced here and in Phase 2, did not create any particular problem for our children. Therefore, the group differences obtained in the first two phases were not due to general short-term memory problems in our poor problem solvers. Other studies (Shafrir & Siegel, 1994; Siegel & Linder, 1984; Siegel & Ryan, 1989) have shown differences in short-term memory between mathematical disabled children and a control group. However, in the latter cases, the selection of the disabled groups was based on difficulties in arithmetic calculation, and they were tested using digit span,

whereas, in our study, the selection was based on problem-solving difficulties and the children (all good in vocabulary ability) were tested with word span.

The check introduced in this phase showed that, in general, the poor problem solvers did not have difficulty in the selection of relevant information in the problems. The data concerning the comparison between groups (where no significant differences were found) are open to different interpretations. Acceptance of the null hypothesis must always be treated with caution, because the testing procedures or statistical tests might be insufficiently discriminative. Since we presented problem types that were very familiar to our children and similar to the problems presented in the preceding phases, we obtained values close to the ceiling for relevant information. This suggests that the task of selecting relevant information was very easy for all of our subjects. For this reason, the most appropriate statistical comparison between the groups concerns the case of irrelevant information. In this case, we found a slightly greater (but not significantly different) tendency in the poor problem solvers to select irrelevant information. This tendency was slight and cannot completely explain, in our opinion, the data obtained in Phase 3 but could have contributed to them.

We should note another possible explanation. Poor problem solvers are able to activate relevant information, but the "level of activation" of this information is not as high as it is for good problem solvers. More precisely, the good problem solvers may have activated the relevant information highly above threshold, whereas the poor problem solvers activate it barely above threshold. If so, it is possible that good and poor problem solvers have the irrelevant information active at the same level. The reason why poor problem solvers perform worse could be because the relevant information is active to a lesser degree than for good problem solvers; therefore, they could have more difficulty discriminating the information (irrelevant vs. relevant), which has similar activation.

Phase 5

The goal of Phase 5 was to verify the persistence of the mathematical disability and of the working memory deficit in our group of poor problem solvers in a follow-up study at the end of the 2-school-year period (the subjects were retested at the end of the fifth grade of primary school).

The subjects were the same as those in the previous phases, although 3 poor problem solvers and 5 good problem solvers were not allowed (by their teachers) to continue the study. Altogether, 13 good problem solvers and 12 poor problem solvers were tested.

Method

Materials and Procedure. The listening test was similar to that used in Phase 1, except that the sentence contents were different and there were two more series of five sentences each. The procedure was the same as that used in the Phase 1. During testing, the teachers were interviewed for information about the problem-solving achievement of their students.

Results

The mean performance of the two groups is presented in Table 5. The data pattern replicates that obtained in Phase 1. The children in both groups did not have any particular difficulty in the verification task (the performance approximated the maximum possible score of 27 for both groups). But all the children (despite some improvements with respect to the preceding year) experienced difficulty in the listening span test, with the mean scores far below the maximum score. The good problem solvers made more correct responses than did the poor problem solvers (3.08 vs. 1.42) [$t(23) = 2.66, p = .014$]. The good problem solvers also correctly recalled more words [$t(23) = 3.14, p = .005$] and recalled more words in perfect order (17.77 vs. 11.75; $p < .002$). The poor problem solvers erroneously remembered three times more words embedded in the sentences than did good problem solvers [$t(23) = 2.64, p = .015$]. The data concerning the subjects' mathematical ability collected from the teachers' interview confirmed all the diagnoses given in the preceding school year.

Discussion

In Phases 1, 2, and 3, we found that the group of poor problem solvers had worse working memory; however, it might be argued that they might have changed over the course of the school year and that the cognitive failures of our poor problem solvers were not due to a persistent problem. Results of this phase reject this possibility. We found that the children diagnosed as poor problem solvers at the beginning of the fourth grade were still included in a mathematical disability group at the end of the fifth grade, whereas the good problem solvers were not. Furthermore, even the observations concerning the listening span test collected in Phase 1 were replicated here, confirming the permanence of the centrality of the effects found in that phase. The successive phase included a similar check for the observations obtained in Phase 2.

Phase 6

The main goal of Phase 6 was to investigate the retention of irrelevant information that had to be processed ini-

Table 5
Mean Performances and Standard Deviations of the Poor Problem Solvers and the Good Problem Solvers in a Working Memory Listening Span Test

Index	Poor Problem Solvers		Good Problem Solvers	
	M	SD	M	SD
T/F	26.42	0.67	26.54	0.52
Correct Rec.	16.17	5.21	21.08	2.06
Intrus.	3.08	2.43	1.15	0.99
Sequen.	1.42	1.44	3.08	1.66

Note—T/F, number of correct true/false responses in the verification of sentences; Correct Rec., number of correctly recalled sentence-final words; Intrus., number of intrusions (i.e., words recalled despite the fact that they were not at the end of a sentence); Sequen., number of sequences of sentence-final words that were completely recalled.

Table 6
Mean Scores, Standard Deviations, and Mean Proportions
(in Parentheses) of Different Items Recalled in the Word Listening
Span Test by the Group of Poor Problem Solvers and the
Group of Good Problem Solvers and the Mean Scores of
the Two Groups on the Recognition Task

	Poor Problem Solvers		Good Problem Solvers	
Index	M	SD	M	SD
Word Listening Span Test				
Sequen.	2.08 (.23)	2.35 (.26)	4.08 (.45)	1.38 (.15)
Correct Rec.	16.08 (.59)	5.01 (.18)	19.31 (.71)	2.35 (.07)
Animal Intrus.	3.17 (.12)	2.76 (.10)	0.92 (.03)	0.86 (.03)
Other Intrus.	2.67 (.03)	2.67 (.03)	1.85 (.02)	1.62 (.01)
Recognition Task				
Mean Score				
False Alarms	2.58	1.56	2.69	2.32
Misses	6.67	2.70	3.85	1.40
Percent Recognition Misses				
Animals	18.75	26.38	7.70	12.01
Nonanimals	47.91	21.36	30.13	12.05

Note—Sequen., proportion of sequences of list-final words that were completely re-called; Correct Rec., proportion of correctly recalled words; Animal Intrus., propor-tion of animal (name) intrusions; Other Intrus., proportion of nonanimal intrusions (i.e., recalled words that were not animal names and were not at the end of the strings).

tially in working memory but did not have to be recalled later. The question was whether this information was com-pletely discharged from the memory system or whether the subjects retained some memory of that information.

We contrasted the two following alternative hypothe-ses: It is possible that good problem solvers do not re-member irrelevant information; conversely, it is possible that good problem solvers remember irrelevant informa-tion but are better able to suppress it when it is necessary to do so than are poor problem solvers.

In order to contrast these two hypotheses, we ran a phase in which we retested the subjects of Phase 5 with a similar version of the listening test used in Phase 2. The task was administered with better controlled materials than in Phase 2 (see below), and a complete set of mate-rials was presented to all subjects. The listening task was followed by an unexpected recognition test, in which the subjects were requested to identify the previously pre-sented items.

Method

Materials. The listening test used strings of animal and nonan-imal nouns and was similar to that used in Phase 2. It used new words based on the norms of Cornoldi and Pra Baldi (1978) in order to have medium–high frequency and a medium–high level of im-agery value. Each string included four two-syllable words. There were 3 sets with 2 strings, 3 sets with 3 strings, 3 sets with 4 strings, for a total of 27 strings and 108 words. Of these words, 26 were an-imal nouns (9 in the final position, and 17 placed in the strings). Likewise, as in Phase 2, the testing phase was preceded by examples.

The recognition list included 32 words: 16 "old" words presented in the listening test, and 16 "new" words used as distractors. The proportion of animal and nonanimal words in the recognition list corresponded to the proportion in the listening test material. In the recognition test, there were 8 animal nouns (4 new animal nouns and 4 old animal nouns, 2 placed in the string and 2 in the final po-

sition of the string) and 24 nonanimal nouns (12 new and 12 old, 6 in the final position and 6 placed in the string).

Procedure. The listening test procedure, with strings of animal and nonanimal nouns, was the same as in Phase 2.

At the end of the listening test, the recognition test was given. The subjects were presented with a recognition list, in which the nouns were written one per row in random order. The children were requested to mark, with a pen, the nouns that they had heard in the previous task, regardless of whether or not the nouns had appeared in the final position. The subjects were asked to be as accurate as possible, and they did not have any time limitation.

Results

Table 6 presents the mean number of correct sequences of final items recalled in the right order, the mean num-ber of correct last items, the mean number of animal in-trusions, and the mean number of other intrusions.

The differences between poor and good problem solvers persisted after the 2-year period. The poor prob-lem solvers remembered fewer completely correct se-quences [$t(23) = 2.61, p = .0157$] and fewer words in the final position [$t(23) = 2.08, p < .05$].

The results concerning the intrusions were similar to those in Phase 2: Poor problem solvers made more intru-sion errors than did good problem solvers, particularly for animal names. A 2 × 2 ANOVA for a mixed design on intrusion errors revealed a significant main effect of the groups [$F(1,23) = 7.47, MS_e = 38.75, p = .012$], a sig-nificant main effect of intrusions type [$F(1,23) = 11.68, MS_e = 27.94, p = .002$] with more animal noun intrusion errors than other noun errors, and a significant interaction type of intrusions × groups [$F(1,23) = 6.50, MS_e = 27.94, p = .018$].

For the correct responses, a larger proportion of animal nouns was recalled than of other nouns: 72% correct re-

call (SD = 16) for animal nouns in the final position and 65% correct recall (SD = 16) of nonanimal nouns in the final position [$t(23)$ =1.97, p = .03].

The two groups' mean performance on the recognition task is presented in the lower part of Table 6. The good problem solvers had fewer misses (i.e., the items presented in the listening test not recognized by the subjects) than did the poor problem solvers [$t(16)$ = 3.39, p < .004]. There was no significant difference between the two groups in the number of false alarms, so we focused the analysis on the number of misses.

A 2 (groups) × 2 (percentage of misses of animal nouns, percentage of misses of nonanimal nouns) mixed ANOVA revealed a main group effect [$F(1,23)$ = 7.25, MS_e = 358.14, p = .013] and a significant main effect of items type [$F(1,23)$ = 24.16, MS_e = 343.85, p < .0001], due to the fact that there were fewer misses of animal nouns in both groups. The interaction between groups and items type was not significant.

Student's t-test comparisons between groups revealed a greater percentage of misses of nonanimal nouns for the poor problem solvers than for the good problem solvers (48% vs. 30%) [$t(23)$ = 2.59, p = .016], but not in the percentage of misses of animal nouns.

Discussion

The results of the working memory test replicated the observations of Phase 2 and showed that the poor problem solvers not only performed worse on the working memory task but also made more intrusion errors. In Phase 6, we also found a significant interaction between groups and type of intrusions, due to the fact that intrusion errors for animal nouns were more frequent in the group of poor problem solvers than in the group of good problem solvers. This result (already observed in poor comprehenders; De Beni et al., 1998) was not observed in Phase 2, perhaps because not all the materials were presented to the subjects in Phase 2. The result implies that intrusion errors are related to the stress given to items. Since items receiving more stress should, in principle, be better remembered (e.g., Craik & Tulving, 1975), the intrusion errors appear to have resulted from an inability to inhibit the retrieval of remembered information, rather than from memory weaknesses. This interpretation of the results is consistent with the observation that, on the recognition test, poor problem solvers did not recognize the embedded items worse than did the good problem solvers.

The final memory test differed from the listening span test both for the delay between presentation and test and for the procedure (recognition vs. recall). Thus, we cannot determine whether the good problem solvers' ability to avoid intrusion errors in the listening span test was due to a reduction of the accessibility of irrelevant information or to an explicit inhibitory effort during the output selection of the items to be retrieved.

GENERAL DISCUSSION

Research has shown that many different cognitive abilities and processes are involved during problem solving (see Mayer, 1992, for a review). This study specifically considered the contribution of working memory to arithmetic word problem solving. We were guided by two main hypotheses: (1) arithmetic word problem-solving ability is related to working memory ability, and (2) this relationship mainly concerns an aspect involved in working memory tasks (i.e., the ability to control irrelevant information).

Analysis of the arithmetic word problem-solving task suggests that the whole working memory system can be involved. A child must perform a series of basic operations to solve a problem presented in either written or oral form (Mayer, 1982, 1992). These operations may concern processing all the material, initial understanding of the problem typology, selection and integration of the critical points, construction of a problem model, planning of the solution pathway, running procedures and calculations, their monitoring, and final evaluation.

The operations require not only a short-term storage function but also more complex active working memory processes. These complex processes are differentiated from the short-term memory functions. These complex processes may be defined as working memory operations in theories that distinguish between short-term memory and working memory (e.g., Swanson, 1994) as examples of active processes in models that assume the existence of a passive–active continuum (Cornoldi, 1995), or they can be defined as belonging to the central executive component in Baddeley's (1986) componential model of working memory.

Our data offer some insight into the nature of a deficit in working memory. In particular, they suggest that a simple storage view (relating this deficit to a lower number of units, which can be maintained in short-term memory) is insufficient to explain why children, with low working memory test scores, have comparable scores on the short-term memory task to the scores of the other children. One hypothesis is that poor problem solvers' storage capacity remains the same, but it is used in a less efficient way. This hypothesis is consistent with literature that connects working memory capacity to the efficiency of inhibition mechanisms (Bjorklund & Harnishfeger, 1990). In fact, poor inhibitory mechanisms (related to the prefrontal functions involved in the central executive component of working memory; Baddeley, 1990) have been proposed in order to explain a variety of failures in control processes (e.g., Shallice, 1988) and have also been associated to failures in text processing (Baddeley, 1990). On the basis of our data, it is not possible to determine whether the inhibition failure is a primary deficit or whether it is connected to other more general factors. In particular, our data are compatible with the

hypothesis of Conway and Engle (1994) and Engle, Conway, Tuholski, and Shisler (1995) that low-span people do not have the attentional resources necessary to inhibit irrelevant information.

Actually, the data from Phases 2 and 6 offer the best evidence of a working memory inhibitory deficit in the poor problem solvers. Indeed, the processing required by the secondary task was presumably very easy and was the same for both groups. Despite this similarity in the secondary task, the poor problem solvers performed worse on the memory task. Furthermore, the secondary task had the effect of increasing some items' probability of being both correctly and incorrectly retrieved. It seems that this emphasis increased the accessibility (Craik & Tulving, 1975) and the difficulty of inhibiting these items. Therefore, we found many intrusions of these items in the listening task, with more intrusions for the poor problem solvers than for the good problem solvers.

It might be questioned whether more intrusion errors indicate a loss of inhibition processes or whether they are related to a selective attention deficit. From a conceptual point of view, it is difficult to disentangle the inhibition and selective attention hypothesis, since selective inhibition can be considered complementary to selective attention. However, the materials that had to be inhibited in our tasks were not completely discharged from the subjects' attention, since they had to be processed for the secondary task. Therefore, the two aspects do not overlap.

Brainerd and Reyna (1988) suggested that most of the results obtained from dual-task experiments could be adequately explained in terms of output interference. Similarly, our data can be interpreted with reference to the retrieval processes and, in particular, to retrieval interference among competing responses. For our dual-task results (Phases 1, 2, 5, and 6), it could be argued that our poor problem solvers had some difficulties not in the encoding of information but in the selection between the competing response in the working memory.

It might also be argued that the failure of the poor problem solvers in the listening tests was due to the fact that they had problems discriminating between different sets of information or maintaining the presentation order. However, this hypothesis cannot explain the two results we obtained—that is, why in other respects the poor problem solvers were as good at memorizing items as the good problem solvers and why, in Phases 2 and 4, the intrusions were more frequent with the animal names, presumably better processed and remembered than the other nontarget names. Furthermore, the dual span condition in Phase 4 showed that the poor problem solvers did not have any more difficulty in handling the secondary task together with the main memory task than did the good problem solvers. The result with the dual span task also offers evidence that poor problem solvers do not necessarily have any more difficulty when they are involved in a more complex task. However, our secondary task was

simple enough that it did not depress the performance with respect to the standard span task. Therefore, we cannot exclude that poor working memory performance and/or inhibitory abilities are related to more demanding situations, perhaps to a lower amount of overall available cognitive resources.

The results from Phases 3 and 4 of this research have helped us to understand the relationship between working memory failure, with the material present in arithmetic word problem solving, and failure with other material. In fact, the data pattern (worse memory for critical information, more memory for irrelevant information for poor problem solvers) was present both in the working memory and in the problem memory tasks.

The results from Phases 5 and 6 indicate that the poor problem solvers showed the same deficit in working memory throughout 2 school years of the research and did not improve their mathematical ability during that period. These data allow us to rule out the possibility that our results might be influenced by confounding variables connected with a maturation effect. Likewise, Geary (1990) found that there were no reliable changes in the mix of strategies or in the rate of execution of counting or in memory retrieval strategies when testing a group of math-disabled children with a specific deficit in arithmetical calculation (defined "no-change group") after a 1-year delay.

REFERENCES

AMORETTI, G., BAZZINI, L., PESCI, A., & REGGIANI, M. (1994). *Test di matematica per la scuola dell'obbligo* [Mathematic test for primary schools]. Florence, Italy: Organizzazioni Speciali.

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

BADDELEY, A. D. (1990). *Human memory: Theory and practice*. Boston: Allyn & Bacon.

BJORKLUND, D., & HARNISHFEGER, K. (1990). The resources construct in cognitive development: Diverse sources of evidence and a theory of inefficient inhibition. *Developmental Review*, **10**, 48-71.

BRAINERD, C. J., & REYNA, V. F. (1988). Generic resources, reconstructive processing and children's mental arithmetic. *Developmental Psychology*, **24**, 324-334.

CONWAY, A. R. A., & ENGLE, R. W. (1994). Working memory and retrieval: A resource dependent inhibition model. *Journal of Experimental Psychology: General*, **123**, 354-373.

CORNOLDI, C. (1995). La memoria di lavoro visuospaziale [Visuospatial working memory]. In F. Marucci (Ed.), *Le immagini mentali* (pp. 145-181). Florence, Italy: La Nuova Italia.

CORNOLDI, C., & COLPO, G. (1981). *Prove di lettura MT* [Reading comprehension test MT]. Florence, Italy: Organizzazioni Speciali.

CORNOLDI, C., DE BENI, R., & PAZZAGLIA, F. (1996). Profiles of reading comprehension difficulties: An analysis of single cases. In C. Cornoldi & J. Oakhill (Eds.), *Reading comprehension difficulties: Processes and intervention* (pp. 113-136). Mahwah, NJ: Erlbaum.

CORNOLDI, C., & PRA BALDI, A. (1978). Il valore d'immagine nei bambini: Norme per 257 nomi e usi nella ricerca applicata [Imagery value in children: Norms for 257 nouns and uses in applied research]. *Formazione e Cambiamento*, **3-4**, 275-304.

CRAIK, F. I. M., & TULVING, E. (1975). Depth of processing and the retention of words in episodic memory. *Journal of Experimental Psychology: General*, **14**, 268-294.

DANEMAN, M., & CARPENTER, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning & Verbal Behavior*, **19**, 450-466.

DANEMAN, M., & CARPENTER, P. A. (1983). Individual differences in integrating information between and within sentences. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **9**, 561-583.

DE BENI, R., PALLADINO, P., PAZZAGLIA, F., & CORNOLDI, C. (1998). The role of inhibition in the working memory deficits of poor readers. *Quarterly Journal of Experimental Psychology*, **51A**, 305-320.

ENGLE, R. W., CONWAY, A. R. A., TUHOLSKI, S. W., & SHISLER, R. J. (1995). A resource account of inhibition. *Psychological Science*, **6**, 122-125.

GEARY, D. C. (1990). A componential analysis of an early learning deficit in mathematics. *Journal of Experimental Child Psychology*, **49**, 363-383.

GERNSBACHER, M. A. (1993). Less skilled readers have less efficient suppression mechanisms. *Psychological Science*, **4**, 294-298.

GERNSBACHER, M. A., & FAUST, M. E. (1991). The mechanism of suppression: A component of general comprehension skill. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **17**, 245-262.

HASHER, L., STOLTZFUS, E. R., ZACKS, R., & RYPMA, B. (1991). Age and inhibition. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **12**, 323-335.

HASHER, L., & ZACKS, R. (1988). Working memory, comprehension, and aging: A review and a new view. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 22, pp. 193-225). San Diego: Academic Press.

HEGARTY, M., MAYER, R. E., & MONK, C. A. (1995). Comprehension of arithmetic word problems: A comparison of successful and unsuccessful problem solvers. *Journal of Educational Psychology*, **87**, 18-32.

HITCH, G. J., & McAULEY, E. (1991). Working memory in children with specific arithmetical learning difficulties. *British Journal of Psychology*, **82**, 375-386.

LEATHER, C. V., & HENRY, L. A. (1994). Working memory span and phonological awareness tasks as predictors of early reading ability. *Journal of Experimental Child Psychology*, **58**, 88-111.

LOGIE, R. H., GILHOOLY, K. J., & WYNN, V. (1994). Counting on working memory in arithmetic problem solving. *Memory & Cognition*, **22**, 395-410.

LOW, R., & OVER, R. (1992). Hierarchical ordering of schematic knowledge relating to area-of-rectangle problems. *Journal of Educational Psychology*, **84**, 62-69.

LOW, R., OVER, R., DOOLAN, L., & MITCHELL, S. (1994). Solution of algebraic word problems following training in identifying necessary and sufficient information within problems. *American Journal of Psychology*, **107**, 423-439.

LUCANGELI, D., & CORNOLDI, C. (1997). Mathematics and metacognition: What is the nature of the relationship? *Mathematical Cognition*, **2**, 121-139.

LUCANGELI, D., & PASSOLUNGHI, M. C. (1995). *Psicologia dell'apprendimento matematico* [Psychology of mathematical learning]. Turin, Italy: Utet.

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

MAYER, R. E. (1992). *Thinking, problem solving, cognition* (2nd ed.). New York: Freeman.

MAYER, R. E., LARKIN, J. H., & KADANE, J. (1984). A cognitive analysis of mathematical problem solving ability. In R. Sternberg (Ed.),

Advances in the psychology of human intelligence (Vol. 2, pp. 231-273). Hillsdale, NJ: Erlbaum.

MEIRAN, N. (1996). Is reading ability related to activation dumping speed? Evidence from immediate repetition priming. *Memory & Cognition*, **24**, 41-59.

MONTAGUE, M. (1992). The effects of cognitive and metacognitive strategy instruction on the mathematical problem solving of middle school students with learning disabilities. *Journal of Learning Disabilities*, **25**, 230-248.

NATHAN, M. J., KINTSCH, W., & YOUNG, E. (1992). A theory of algebra word problem comprehension and its implications for the design of learning environments. *Cognition & Instructions*, **4**, 329-390.

OAKHILL, K., & YUILL, N. (1996). Higher order factors in comprehension disability: Processes and remediation. In C. Cornoldi & J. Oakhill (Eds.), *Reading comprehension difficulties: Processes and intervention* (pp. 69-92). Mahwah, NJ: Erlbaum.

PASSOLUNGHI, M. C., CORNOLDI, C., & DE LIBERTO, S. (1996, July). *Influence of working memory on word problem solving in children.* Paper presented at the Second International Conference on Memory, Abano-Padua, Italy.

PASSOLUNGHI, M. C., LONCIARI, I., & CORNOLDI, C. (1995, August). *The effects of metacognition categorization and planning ability on learning problem solving.* Paper presented at the 6th Early European Conference, Nijmegen, The Netherlands.

PRESSLEY, M. (1990). *Cognitive strategy instruction that really improves children's academic performance.* Cambridge, MA: Brookline.

REYNOLDS, C., & BIEGLER, E. (1987). *Test of memory and learning.* Austin, TX: ProEd.

SHAFRIR, U., & SIEGEL, L. S. (1994). Subtypes of learning disabilities in adolescents and adults. *Journal of Learning Disabilities*, **27**, 123-134.

SHALLICE, T. (1988). *From neuropsychology to mental structure.* Cambridge: Cambridge University Press.

SIEGEL, L. S., & LINDER, B. A. (1984). Short-term memory processes in children with reading and arithmetic learning disabilities. *Developmental Psychology*, **20**, 200-207.

SIEGEL, L. S., & RYAN, E. B. (1988). Development of grammatical-sensitivity, phonological, and short-term memory skills in normally achieving and learning disabled children. *Developmental Psychology*, **24**, 28-37.

SIEGEL, L. S., & RYAN, E. B. (1989). The development of working memory in normally achieving and subtypes of learning disabled children. *Child Development*, **60**, 973-980.

SWANSON, H. L. (1994). Short-term memory and working memory: Do both contribute to our understanding of academic achievement in children and adults with learning disabilities? *Journal of Learning Disabilities*, **27**, 34-50.

SWANSON, H. L., COONEY, J. B., & BROCK, S. (1993). The influence of working memory and classification ability on children's word problem solution. *Journal of Experimental Child Psychology*, **55**, 374-395.

THURSTONE, L. L., & THURSTONE, T. G. (1941). *Factorial studies of intelligence* (Psychometric Monographs, No. 2). Chicago: University of Chicago Press.

TURNER, M. L., & ENGLE, R. W. (1989). Is working memory capacity task dependent? *Journal of Memory & Language*, **28**, 127-154.