

## The Role of Conceptual Knowledge and Metamemory in the Development of Organizational Processes in Memory

WOLFGANG SCHNEIDER

*Max Planck Institute for Psychological Research*

The present study investigated the relationship between developmental shifts in the organization of materials and developmental changes in deliberate strategy use. Second and fourth grade children were presented with clusterable sort/recall lists representing the factorial combinations of high and low interitem association, and high and low category relatedness. Strategy use in the task was rated by the experimenter and also assessed via self reports. General and task-related strategy knowledge (metamemory) was also examined. Second graders displayed more category clustering during recall for highly associated items than for weakly associated items, whereas older children's recall organization (but not recall) was unaffected by this organizational dimension. Correlations among measures of metamemory and organizational behavior indicated that second graders in general were unaware of the importance of categorization strategies for facilitation of recall. On the other hand, sorting during study and task-related metamemory were the most important predictors of fourth graders' recall performance, thus indicating that most fourth graders used categorization strategies deliberately.

© 1986 Academic Press, Inc.

Over the past 10 years, a major focus in memory development research has concerned the development of organization in children's memory. Age differences in categorization and subsequent memory performance in free recall and sort/recall tasks have been explained as increasing sophistication in the use of organizational strategies, mainly sorting during study and clustering during recall (Lange, 1978; Moely, 1977; Ornstein & Corsale, 1979). According to this viewpoint, organization in recall is the result of strategic processes that represent deliberate and effortful attempts to encode and retrieve information.

The research presented in this paper was prepared during a stay at the Department of Psychology, Stanford University, which was made possible by a grant from the Stiftung Volkswagenwerk, Hannover, Federal Republic of Germany (AZ.: II/37 129). I am grateful to John Flavell for his support and many hours of discussion. Thanks are further due to Christel Uhl for her assistance in conducting the study, and to Merry Bullock for helpful comments on an earlier version of the manuscript. Reprint requests should be sent to Wolfgang Schneider, Max-Planck-Institut für psychologische Forschung, Leopoldstrasse 24, D-8000 Munich 40, West Germany.

The validity of this interpretation has been debated for various reasons. First, the definition of memory strategies as deliberate and voluntary is questioned because automatic, unconscious activation characterizes the use of highly learned strategies (Naus & Ornstein, 1983; Pressley, Forrest-Pressley, Elliott-Faust, & Miller, 1985). Moreover, even if the traditional definition of a memory strategy is accepted, there is strong empirical evidence that mnemonic context, that is, situational factors such as the quality of instructions, influences the deployment of mnemonic strategies in free-recall or sort/recall tasks (Corsale & Ornstein, 1980). In addition to instructional effects, characteristics of the to-be-remembered materials such as the perceived typicality of category examples and the degree of associativity among categorically related items also influence young children's memory behavior. For example, clustering is observed in kindergarten and young elementary school children when the category structure is made highly salient to them, that is, when category typicality or associativity among categorically related items is relatively high (see for a review, Bjorklund, 1985).

However, it has not been clearly demonstrated that such categorization during sorting or recall reflects the deliberate utilization of a mnemonic strategy. Lange (1973, 1978), for example, has argued that young children's clustering during free recall may be automatically determined by the associative structure of the task materials. Two recent studies (Bjorklund & de Marchena, 1984; Frankel & Rollins, 1985) presented empirical evidence supporting the position that a developmental shift from memory organization based on associative criteria to memory organization based on taxonomic criteria can be observed during the elementary school years and may underlie age-related organization in memory tasks. Bjorklund and de Marchena (1984) demonstrated that older children (fourth and seventh graders) organized items according to taxonomic categories and did not use associative relations in their sorts, whereas first grade children showed a tendency toward associative organization when the list structure allowed both associative and categorical organization. Empirical support for Lange's position includes a study by Frankel and Rollins (1985) who assessed the impact of associative versus categorical relations by factorially manipulating high and low category relatedness and high and low interitem associations. They found that fourth and tenth graders showed relatively high levels of organization in recall whenever category relatedness or associative strength were high. In contrast, kindergarteners displayed greater category clustering only for the list compatible with associative organization (i.e., taxonomic factors did not affect performance).

Taken together, these results suggest that younger children's organization during study as well as during recall may be guided by automatic associations between items rather than by deliberate category grouping strategies, and thus support Lange's (1973, 1978) theoretical claim that

relatively automatic associative organization principles mediate young children's recall and output organization. Additionally, Bjorklund and de Marchena (1984) provided evidence that young children's clustering during study (in contrast to recall) may be due to automatically activated associative relations. According to the authors, the shift from greater associative organization in first graders to greater categorical organization in fourth and seventh graders does not reflect a shift in the deliberate use of two memory strategies but rather a shift in the ease with which semantic relationships are activated. That is, age changes in memory organization observed through the elementary school years may be attributable primarily to developmental changes in children's semantic memory or knowledge base and not to the emergence of deliberate memory strategies (see also Bjorklund, 1985).

One problem with this interpretation is that no in-depth analysis of semantic memory or the knowledge base is available (Ornstein & Naus, 1985). Moreover, several conceptualizations of semantic memory can be found in the literature (cf. Nelson & Brown, 1978). The term is typically equated with "world knowledge", including anything the child has learned (e.g., Chechile & Richman, 1982). As such, knowledge base is a much broader conceptualization than strategic or metamemory considerations. This makes it particularly difficult to assess the mechanisms by which the knowledge base actually mediates memory performance. As a consequence, Ornstein and Naus (1985) emphasize the need for data concerning what children of different ages know about specific content domains.

In addition to the role of semantic organization, children's strategic behavior may be related to their metamemory, that is, to their knowledge about memory processes and strategies. This is an area of knowledge that has been shown to be relevant in many memory situations in that it facilitates strategic behavior (Flavell & Wellman, 1977; Schneider, 1985). As there is reason to assume that developmental changes in children's metamemory are related to their use of memory strategies, an experiment designed to provide a test of the role of deliberate memory strategies in the development of organizational behavior should consider possible influences of metamemorial knowledge.

Thus the goal of the present study was to compare the relative effects of task-inherent organization (i.e., semantic memory, cf. Bjorklund, 1985) and metamemory on second and fourth graders' memory behavior and performance in a sort/recall task. In order to assess the effects of semantic memory, the item lists originally developed by Frankel and Rollins (1985) were used in this study. That is, four sets of items to be recalled were created by crossing high and low category relatedness with high and low interitem association. Subjects were presented with one of these sets and instructed to do anything with the items that would help them to

remember them better. To assess deliberate strategy use, we included two modifications or "enrichments" of the traditional sort/recall task measures:

(1) A distinction was made between sorting and studying. Theoretically, enormous recall differences can be expected for subjects who show identical sorting behavior but differ with regard to the way they study the items. For instance, better recall performance would be expected for subjects who use more sophisticated study strategies like cumulative rehearsal and self-testing after complete categorization of the items. Thus, both sorting and studying behavior were recorded separately in the present study.

(2) An attempt was made to assess subject's verbalizable knowledge about memory processes and strategies relevant for the solution of the sort/recall task in question (i.e., task-related metamemory) as well as their knowledge about different aspects of prospective and retrospective memory strategies relevant for memory in school-related and everyday life contexts (i.e., general metamemory). Because of conceptual and methodological requirements, a multimethod assessment was made in this study (see for a detailed discussion, Cavanaugh & Perlmutter, 1982; Schneider, 1985). This included an interview similar to that developed by Kreutzer, Leonard, and Flavell (1975) which taps general metamemory, an interview dealing with task-related metamemory, and a paired comparison judgment task developed by Justice (1986) to assess children's judgments of strategy effectiveness.

## METHOD

### *Subjects*

The subjects were 128 middle-class children, 64 each from the second and fourth grades. Children were selected from two public elementary schools located in the Rhein-Neckar area near Heidelberg, West Germany. Mean ages of the two groups were 7.10 years (range 7.3–8.4) and 10.2 years (range 9.5–10.8). An equal number of boys and girls was included at each age level.

### *Materials*

*Sort-recall task.* Free recall lists were generated by crossing two levels (high and low) of relatedness between stimulus items and category labels by two levels (high and low) of associations among individual items within each category. Sixteen subjects (eight males and eight females) of each grade level were randomly assigned to one of the four sort/recall lists (described below). The stimuli were four sets of six black and white line drawings of common objects adapted from Frankel and Rollins (1985). Each picture was drawn on a 4.2 × 4.2 cm card. The pictures in each list are described in Table 1.

TABLE 1  
 DESCRIPTIONS OF THE FOUR SETS OF PICTURE STIMULI USED IN THE SORT/RECALL TASK

List A: High category relatedness, high interitem associativity

*Animals:* dog, cat, horse, cow, pig, mouse  
*Clothing:* jacket, socks, pants, shoe, coat, hat  
*Vehicles:* car, bus, airplane, train, truck, bicycle  
*Furniture:* chair, table, bed, sofa, desk, lamp

List B: High category relatedness, low interitem associativity

*Animals:* tiger, elephant, cow, pig, bear, dog  
*Clothing:* coat, shoe, dress, pants, sweater, tie  
*Body parts:* hand, stomach, eye, toes, leg, ear  
*House parts:* window, door, chimney, floor, roof, brick

List C: Low category relatedness, high interitem associativity

*Animals:* goat, deer, hippopotamus, buffalo, monkey, lamb  
*Clothing:* belt, gloves, scarf, vest, suit, sweater  
*Body parts:* ankle, knee, elbow, neck, back, shoulder  
*House parts:* patio, porch, fence, gate, steps, railing

List D: Low category relatedness, low interitem associativity

*Animals:* beaver, rat, alligator, camel, squirrel, giraffe  
*Clothing:* belt, gloves, scarf, vest, suit, sweater  
*Vehicles:* tractor, wagon, helicopter, tank, sled, rocket  
*Furniture:* refrigerator, stool, rocker, bookcase, stove, bench

Interitem association values were derived from the Marshall and Cofer (1970) and Palermo and Jenkins (1964) norms. Category relations were maximized or minimized according to the norms of Battig and Montague (1969). Although equivalent norms of prototypicality and interitem associativity do not exist in Germany, an inspection of German typicality norms (Flammer, Burkhardt, Jann, Reisbeck, and Stadler, 1985) revealed that they were roughly comparable, e.g., all items defined as highly related to the category by Frankel and Rollins (1985) ranked in the top 25% of category exemplars according to Flammer et al. (1985), and nearly all of the items showing low category relatedness according to Frankel and Rollins also ranked near the bottom of the German norm lists (the only exception was the item "jacket" which ranked 10th in the Flammer et al. norm list).

In addition, a procedure similar to that used by Frankel and Rollins (1985) was chosen to validate categories and associations used in the present study. An independent sample of 25 second graders was selected, and asked to sort the items. With the exception of two subjects, these children correctly grouped all the test materials into categories. As in the Frankel and Rollins study, children were further asked to choose first, second, and third choice pairs of highly associated items within

each category. It was expected that normed associates should be picked together more often than with other items in the high associate lists and that greater dispersion in pair choices should occur in the weakly associated sets. Similar to the findings by Frankel and Rollins, these expectations were confirmed for the German sample. Taken together, these findings justified the use of translations of the Frankel and Rollins list.

*Assessment of metamemory.* All subjects were presented with a metamemory questionnaire and a strategy judgment task. The metamemory questionnaire included 11 items concerning planful behavior for a future event, recognizing task difficulty, organized memory search for past events, and strategies for remembering. There were two parts. The first, assessing general metamemory, included six items. Five were expanded versions of questions from the Kreutzer et al. (1975) questionnaire: Preparation Object, Retrieval Event, (two items), and Rote Paraphrase (two items). These questions have been shown to be the most valid and sensitive indicators of metamemory in the Kreutzer et al. questionnaire for children between 6 and 9 years of age (cf. Cavanaugh & Borkowski, 1980; Kurtz & Borkowski, 1984). The sixth item assessed subjects' knowledge about the importance of retrieval cues when memorizing prose materials.

The second, more task-related part of the metamemory questionnaire included five items, each of which required the subjects to compare the difficulty of two same length word lists, one containing clusterable, the other nonclusterable items. The word list pairs differed with regard to list length (between 6 and 15 words) and the arrangement of clusterable items within the lists (i.e., blocked or random order). The five items were given in a fixed sequence, with the more difficult comparisons (i.e., those tasks including longer word lists and random ordering of clusterable items within the clusterable word list) always preceding those comparisons where the differences between the two lists were more obvious. The items in the lists were different from those used in the sort/recall task.

The second task was designed to provide more detailed information about subjects' knowledge of strategies useful for a sort/recall task. We prepared black and white videotapes featuring a 7-year-old male model who demonstrated four different strategies. The videotaped activities were filmed from behind the model, over his shoulder to minimize cues as to his sex. The videotape began with a short introductory period, and then presented four 30-s demonstrations of each of four memory strategies: (a) Grouping: the actor placed 12 pictures into a  $3 \times 4$  matrix of four categories and named aloud each group of three together, twice, e.g., *zebra, elephant, fox, zebra, elephant, fox*. (b) Rehearsal: the actor placed 12 pictures into a random  $3 \times 4$  matrix and named each row twice, e.g., *bus, shoes, hammer, bus, shoes, hammer*. (c) Naming: the actor labeled 12 pictures (in a random matrix) one at a time with no spatial rearrangement.

(d) Looking: the actor visually focused on the pictures one at a time with no spatial rearrangement. Each strategy was named on the videotape.

After watching the videotape, subjects were given a paired comparison task. The names of the demonstrated strategies were written on  $3.5 \times 7.5$  cm cards and combined into 12 pairs, made up of all possible permutations of the four strategies. For each pair, children were asked to state which strategy was better for free recall of a picture list. The 12 trials were divided into two blocks of six trials presented in counterbalanced order to half of the children at each age level. After the 12 paired comparison trials, children were also asked to rank order the four strategies according to their perceived efficacy.

*Measures of intelligence and memory capacity.* Indicators of verbal and nonverbal intelligence were used to control the influence of intelligence and verbal comprehension on metamemory. Two subtests of the Culture Fair Intelligence Test (Cattell & Weiss, 1977) were used to assess nonverbal intelligence. The "Classification" subtest required children to mark the object or geometric pattern in a row of objects or geometric patterns that did not "go together" with the other ones. The "Matrices" subtest consisted of 12 geometric patterns with one part missing in each item. The child was instructed to identify which of five alternatives correctly completed the geometric pattern. Different item sets were used for second and fourth graders.

The vocabulary subtest of the General School Achievement Test (Fip-pinger & Rieder, 1978) containing 19 items was selected to assess children's verbal ability. Each key word was underlined. Subjects selected a word out of five possible choices that corresponded most closely to the key word. Raw scores were used for the two nonverbal intelligence subtests as well as for the verbal ability measure.

Measures of short-term and long-term memory capacity were included to allow for a more accurate estimate of strategy impact on recall (it should be noted that the term "memory capacity" is used in the sense that the measures represent memory functioning most likely astrategic in nature). A slightly modified version of the WISC digit span test was used to assess subjects' short-term memory capacity. In addition, a second measure was constructed to test long-term memory capacity. A list of 20 unrelated items (e.g., cake, bell, eye, drum, castle, etc.) was read to the subjects three times, in a varied order to minimize the effects of strategy use. Children were instructed that they had to reproduce as many items as possible about half an hour later.

### *Procedure*

Each subject participated in two 45-min sessions. In the first session, subjects were presented with the measures of verbal and nonverbal intelligence and the short-term memory measure. Testing was conducted within the classroom in groups of 15 to 20. During the second session,

each child was tested individually. Children were told that they would be shown a set of pictures and should try to do anything that would help them remember the items of the stimulus set. They were then given a 43 × 30 cm metal board with one of the sets of 24 picture cards described in Table 1. The pictures were randomly ordered in four rows of six items. Subjects were asked to name all the pictures, and the experimenter mentioned their categorical relationship. This was done to ensure that category knowledge and the use of category knowledge were not confounded. A 2-min sorting period was allowed where children were given an opportunity to rearrange items in order to learn them most efficiently.

Next, 2 more min were given for studying the items. Children's behavior during sorting and studying was rated, and subjects' arrangement of picture cards was photographed. In addition, after recall the subjects were asked to describe what they had done in order to learn the items so that we could compare children's descriptions with ratings of their sorting/study behavior. Correlations between rated sorting behavior and the index of sorting derived from the photographs on the one hand and correlations between experimenter ratings of studying behavior and children's self reports concerning their study strategies on the other hand were generally high (ranging between .54 and .91). Consequently, the rating variables were felt to exhibit sufficient validity to use them in further analyses.

After the study period, the board and picture cards were removed, and recall was requested. Following the sort/recall task, the long-term memory capacity task was given. Subjects were instructed that a list of 20 unrelated words would be read to them three times, and that they should try to remember the items at the end of the session (i.e., 30 min later). Immediately after the three presentations of the word list, subjects were shown the memory strategy videotapes. They were given the 12 paired comparison trials, and then asked to rank-order the strategies. Next, the metamemory interview was given. At the end of the session, subjects were asked: "We still have to complete one task. Do you know which I mean?" With the exception of two fourth graders, all subjects had forgotten about the long-term memory capacity recall test, which was given next. If the child was not able to remember any item, the first two words from the first reading of the list were presented as retrieval cues (five second graders and two fourth graders were cued on the long-term memory test.) If the child did not recall any words for 1 min, the session was terminated.

## RESULTS

The data to be presented were initially examined for sex of subject. Since sex had no significant effects, the data were collapsed across this variable in all subsequent analyses.



*Recall.* A Grade (2)  $\times$  Interitem Associativity (2)  $\times$  Category Relatedness (2) factorial analysis of variance on the recall data yielded main effects for grade,  $F(1, 120) = 23.74, p < .01$ , Interitem Associativity ( $F(1, 120) = 8.19, p < .01$ ), and Category Relatedness ( $F(1, 120) = 5.06, p < .05$ ). These main effects were qualified by an interaction between Grade and Interitem Associativity,  $F(1, 120) = 4.13, p < .05$ . Interitem associativity affected fourth graders' recall (high associates,  $M = 16.34$ ; low associates,  $M = 12.84, t(62) = 3.06, p < .01$ ), but not that of second graders (high associates,  $M = 11.40$ ; low associates,  $M = 10.81, F(62) = 0.66, n.s.$ ). These findings differed from those obtained by Frankel and Rollins (1985) in that no interaction between category relatedness and interitem associativity was found.

*Organization during sorting and recall.* The adjusted ratio of clustering (ARC) score (Roenker, Thompson, & Brown, 1971) was used as a measure of both sorting during study and clustering during recall. This measure has been shown to be relatively independent of the absolute level of recall (Murphy, 1979). As all subjects who manipulated the items sorted them by row, there was no problem with using the adjusted ratio of clustering score as an index of sorting behavior.

Table 2 contains the clustering scores for organization during study and during recall as a function of Grade and Stimulus Condition. These data were analyzed in two Grade (2)  $\times$  Category Relationship (2)  $\times$  Interitem Association (2) factorial analyses of variance with the two clustering measures as dependent variables. With regard to sorting during study, only a significant effect of Grade,  $F(1, 120) = 6.56, p < .01$ , was found. On average fourth graders employed more categorical sorting (ARC = .52) than second graders (ARC = .29).

With regard to clustering during recall, significant main effects were found for Grade,  $F(1, 120) = 17.78, p < .01$ , and for Interitem Associativity,

TABLE 2  
MEAN RECALL, CATEGORY CLUSTERING DURING SORTING AND DURING RECALL AS A  
FUNCTION OF GRADE, CATEGORY RELATEDNESS, AND INTERITEM ASSOCIATION

Grade and task	High associativity		Low associativity	
	High relatedness	Low relatedness	High relatedness	Low relatedness
Grade 2				
Recall	12.37	10.44	10.75	10.88
Clustering/sorting	.38	.32	.35	.13
Clustering/recall	.63	.38	.18	.16
Grade 4				
Recall	17.31	15.37	14.19	11.50
Clustering/sorting	.70	.77	.40	.20
Clustering/recall	.70	.57	.56	.52

$F(1, 120) = 13.93, p < .01$ . The main effect for Category Relatedness was marginally significant,  $F(1, 120) = 3.69, p < .06$ . A two-way interaction between Grade and Interitem Association,  $F(1, 120) = 4.13, p < .05$ , indicated that low interitem association strength had a large negative impact on younger subjects' clustering scores (high associates,  $M = .51$ , low associates,  $M = .17, t(62) = 3.92, p < .01$ ). The effect of interitem association strength was not significant for fourth graders (high associates,  $M = .64$ ; low-associates,  $M = .54, t(62) = 1.23, n.s.$ ).

As has been pointed out by Frankel and Rollins (1982), deliberate strategy use might be reflected by significant correlations between clustering during sorting and clustering during recall on the one hand and between clustering and recall on the other hand. Clustering during sorting, collapsed across experimental conditions, was correlated with clustering during recall and recall within each age group. The six correlation coefficients (all significant at  $p < .01$ ) obtained were .30 (Clustering during Study and Recall), .35 (Clustering during Study and Clustering during Recall), .48 (Clustering during Recall and Recall) for second graders and .66 (Clustering during Study and Recall), .63 (Clustering during Study and Clustering during Recall), and .53 (Clustering during Recall and Recall) for fourth graders. These findings resemble those of Kee and Bell (1981) who emphasized that organization during recall was not a source of developmental differences in free recall in their study, whereas organization at study increased with age. On the other hand, our results differ from those obtained by Frankel and Rollins (1985) who found significant intercorrelations only for the oldest age group (i.e. tenth graders).

*The impact of the distinction between sorting and studying.* Experimenter ratings of subjects' sorting and studying behavior were used to test the effects of sorting and studying on subsequent recall. Not surprisingly, it was found that, regardless of age, children with high scores in sorting and studying (i.e., with at least 3 out of 4 points in both rating scales) recalled significantly more than the rest of their age group. The mean recall of the more strategic group for second graders was 12.85, compared to 9.25 for the less strategic group,  $t(64) = 3.52, p < .01$ , and for fourth graders, 17.55 compared to 12.52,  $t(64) = 4.87, p < .01$ . However, the hypothesis that recall would also differ between children with comparable sorting scores but differing studying behaviors was only partially supported by the data. Contrary to expectations, subjects who showed perfect or almost perfect sorting also recalled more than did the average child in their age group, with differences in observed study behavior having no significant impact on recall performance. Similarly, fourth graders who did not sort the items performed poorly in recall independent of the rated quality of study behavior. On the other hand, planful study behavior (i.e., use of strategies like rehearsal and self-testing) seemed to affect recall performance of second graders with low sorting scores although

the mean differences (13.25 vs 10.09) did not prove to be statistically significant.

*Assessment of general metamemory and task-specific strategy knowledge.* Each item on the metamemory interview was given a score of 1 or 0. Separate scores were computed for the six general metamemory items (maximum score of 6) and five task-related or specific metamemory items (maximum score of 5). These metamemory data were analyzed separately in two Grade (2)  $\times$  Category Relationship (2)  $\times$  Interitem Association (2) factorial analyses of variance. In both analyses, only an effect of Grade was found,  $F(1, 118) = 9.36, p < .01$ , and  $F(1, 118) = 11.34, p < .01$  for general and specific metamemory scores, respectively. Hence, the speculation that the nature of the stimulus materials in the sort/recall task would affect the metamemory performance that followed it was not confirmed. The mean metamemory scores are listed in Table 3.

Judgments on the paired comparison task were analyzed in terms of the proportion of times each strategy was chosen over every other strategy. About 60% of the second graders and 75% of the fourth graders consistently judged one strategy as superior to the other across the paired judgment presentations, a level which is sufficiently reliable to warrant such a paired comparison analysis (Thurstone, 1927). The resulting proportions (listed in Table 3) reflect the order of the strategies along a continuum of judged effectiveness for a picture recall task. A test for the difference between correlated proportions ( $p < .05$ , corrected for continuity, McNemar, 1969) indicated that although second graders were less likely to choose the labeling strategy than any of the other three strategies, there was no significant preference among the remaining three strategies. In contrast, fourth graders exhibited a marked preference for the repetition

TABLE 3

MEANS AND STANDARD DEVIATIONS (IN PARENTHESES) FOR THE TWO METAMEMORY COMPONENTS, THE RANK-ORDERING PROCEDURE, AND THE SCALE VALUES OF THE PAIRED COMPARISON ANALYSIS, SEPARATELY FOR EACH GRADE

Measure	Grade	
	2	4
General metamemory	4.48 (1.16)	5.05 (0.85)
Task-related metamemory	2.36 (1.45)	3.23 (1.45)
Paired comparison judgments		
Grouping	.425	.771
Repeating	.341	.642
Looking	.234	.005
Labeling	.000	.000
Rank ordering of strategies	1.28 (1.31)	1.97 (1.02)

and grouping strategies. For the fourth graders, there were no differences between grouping vs repeating on the one hand (both were highly preferred) and looking vs labeling (neither was preferred). Although the ordering of strategies in the paired comparison analysis was the same for both age groups, the magnitude of the differences between the scale values differed considerably across grades.

Subjects' rank orderings of the four strategies were compared to adult judgments. Three points were given for the rank order preferred by adults (i.e., Grouping—Repeating—Labeling—Looking). Two points were given for rank orders differing from this "optimal" sequence in one aspect, one point for a sequence that differed in at least two aspects and so on (see Table 3 for the summary statistics on this assessment). As expected, significant age differences ( $p < .01$ ) were found in the extent to which rank orderings matched those of adults. The fourth graders' rankings were more similar to the adult order ( $M = 1.97$ ) than were second graders' ( $M = 1.28$ ),  $t(126) = 2.92$ ,  $p < .01$ .

When the different measures of declarative metamemory (i.e., general and task-specific metamemory) and perceived strategy effectiveness were intercorrelated, almost no significant correlations could be found for the second-graders (the only exception was the correlation between the individual preference for the grouping strategy and the rank ordering procedure,  $r(64) = .58$ ,  $p < .01$ ). On the other hand, generally significant but moderate intercorrelations among the metamemory and strategy effectiveness measures were found for the fourth graders, with correlations ranging from .28 to .34.

In sum, these data indicate that signs of metamemorial awareness can be observed in the older but not in the younger subjects. To assess the effects of metamemory on study and recall, within the fourth graders, three additional analyses were performed. We divided fourth graders into two groups, based on metamemory scores: Subjects in the high metamemory group had scores of at least 4 (out of 5) in the task-specific metamemory interview, and preferred grouping strategies over the three remaining strategies in the paired comparison judgment task. According to these classification criteria, 28 of the 64 fourth graders were selected for the high metamemory group. We then performed three separate analyses of variance on recall, study organization, and recall organization, with high/low metamemory, category relatedness, and interitem associativity as independent variables. The analysis for the recall data yielded main effects for metamemory,  $F(1, 63) = 5.79$ ,  $p < .05$ , category relatedness,  $F(1, 63) = 4.48$ ,  $p < .05$ , and interitem associativity,  $F(1, 63) = 10.27$ ,  $p < .01$ . Mean recall of the high metamemory group was higher than recall of the low metamemory group (16.41 vs 13.64), and recall was higher for lists with high category relatedness (15.75 vs 13.43) and lists with high interitem associativity (16.34 vs 12.85). The interaction among

metamemory, category relatedness, and interitem associativity approached significance,  $F(1, 63) = 3.24, p < .10$ ), showing that mean recall of high metamemory subjects tended to be low only when both category relatedness and interitem associativity were low.

Somewhat surprisingly, study organization was only influenced by interitem association,  $F(1, 63) = 11.56, p < .01$ . Mean cluster scores were higher for lists with high interitem associativity, compared to lists with low interitem associativity (.73 vs .30). No significant main effects were found in the analysis using clustering during recall as the dependent variable. From these data, it can be concluded that aspects of both metamemory and semantic memory affect fourth graders' recall, but that their impact on clustering during study and clustering during recall is less pronounced.

In order to explore the effect of task-specific metamemory on fourth graders' clustering during study, clustering during recall, and recall more thoroughly, a series of multiple regressions was conducted. Instead of dichotomized metamemory scores, raw scores were chosen to make use of all information in the variable. These regression analyses showed that metamemory scores predicted clustering during study but not clustering during recall. To test the assumption that metamemory influences recall through a mediating effect on study organization, a multiple regression with recall as the dependent variable and metamemory and study organization as predictor variables was conducted. The hypothesis was that if metamemory has only a mediating effect on study organization, it should not account for unique variance relative to study organization in recall. However, the results of the multiple regression analysis indicated that metamemory as well as study organization significantly contributed to predicting fourth graders' recall (the standardized regression coefficients for metamemory and study organization were .28 and .50, respectively). It thus appears that metamemory has a direct as well as an indirect impact on older subjects' recall.

*The role of memory capacity, metamemory, and organization in predicting recall.* In order to assess the relative impact of metamemory, memory organization, intelligence, and memory capacity (collapsed across experimental conditions) on children's performance in the sort/recall task, multiple regression analyses were run separately for each grade. As can be seen from Table 4, different patterns of results were obtained for the second and fourth graders. Memory capacity and clustering during recall were the best predictors of second graders' recall performance, whereas metamemory and sorting behavior did not significantly contribute to the prediction of recall. In contrast, sorting during study and task-related metamemory were the most important predictors of fourth graders' recall performance. Interestingly, for both second and fourth graders individual differences in nonverbal and verbal intelligence were not related

TABLE 4  
 MULTIPLE REGRESSION OF METAMEMORY MEASURES, ORGANIZATION MEASURES, AND  
 MEMORY CAPACITY MEASURES ON RECALL

Grade and task	Beta	F	p
Grade 2			
Task-related metamemory	.01	0.01	.93
General metamemory	.18	2.96	.09
Sorting during study	.13	1.43	.24
Clustering during recall	.22	3.94	.04
Nonverbal intelligence	.09	0.58	.45
Vocabulary	.14	1.53	.22
Short-term memory capacity	.33	10.13	.01
Long-term memory capacity	.42	14.48	.01
Grade 4			
Task-related metamemory	.22	5.68	.02
General metamemory	.10	1.29	.26
Sorting during study	.44	15.96	.01
Clustering during recall	.20	3.21	.08
Nonverbal intelligence	.14	2.23	.14
Vocabulary	.08	0.60	.44
Short-term memory capacity	.19	3.24	.08
Long-term memory capacity	.06	0.37	.54

to recall. Both regression models explained a large amount of variance in the dependent variable, with multiple *R* squares for second graders and fourth graders of .49 and .63, respectively.

To explore the importance of memory capacity, memory behavior, and metamemory in explaining age differences in recall, an analysis of covariance on recall with age group as the independent variable and the two measures of memory capacity, the two adjusted ratio of clustering scores reflecting sorting during study and clustering during recall, and the two components of metamemory as covariates was conducted.

The main result was that age differences in recall (second graders:  $M = 11.32$ , fourth graders:  $M = 14.55$ ,  $t(126) = 4.33$ ,  $p < .01$ ) were eliminated after adjustment was made for the linear effects of the covariates (adjusted means: 12.59 and 13.18 for second and fourth graders, respectively;  $t(126) = 0.35$ , n.s.). With the exception of general metamemory, all other covariates proved to be important predictors of age differences in recall.

One possible problem with the interpretation of these results concerns the use of memory capacity as a covariate. If the memory capacity measures indicate only how well subjects did on tasks very similar to the sort/recall procedure, it would not be surprising that using it as a covariate would remove the age effect in the sort/recall procedure. However, the empirical findings do not support this view. First, no substantial

age effects were obtained for the two memory capacity measures. Although fourth graders recalled significantly more items ( $M = 5.32$ ) in the digit-span measure than second graders ( $M = 4.78$ ),  $t(126) = 3.28$ ,  $p < .01$ , no significant age differences were found for the long-term memory capacity measure,  $t(126) = 1.85$ ,  $p > .05$ . Here, fourth graders' mean recall was only slightly higher than that of second graders (4.08 vs 3.72). Second, the intercorrelations among recall in the sort/recall task and recall in the two memory capacity measures were not particularly high. The four coefficients obtained were .29 (sort/recall and digit span), .50 (sort/recall and long-term memory capacity) for second graders and .46 (sort/recall and digit-span) and .31 (sort/recall and long-term memory capacity) for fourth graders.

In order to rule out the possibility that the long-term memory capacity measure did not provide just one additional test of strategy use rather than tap capacity, intercorrelations among this measure, the two metamemory scores, study and recall clustering were also computed, separately for the two age groups. All resulting correlation coefficients were insignificant, ranging between  $-.07$  and  $.19$ . In view of these correlations the finding that memory capacity turned out to be an important predictor of age differences in recall does not seem trivial.

## DISCUSSION

The main purpose of the modifications and enhancements of the traditional sort/recall task in the present study was to provide a more thorough test of the hypothesis that the shift from more associative organization to a more categorical organization in older elementary school children does not reflect a shift in the deliberate use of memory strategies but a shift in the ease with which different semantic relationships are activated (see Bjorklund, 1985). As developmental shifts in memory behavior in sort/recall tasks may not only be attributable to developmental differences in semantic memory but also to qualitative and quantitative differences in children's metamemory (cf. Schneider, 1985), measures of metamemory were also used to assess the role of deliberate strategy use in elementary school children's memory behavior. An explicit category cue for each stimulus item insured that category knowledge and the use of this knowledge were not confounded. However, very few second graders (about 10%) spontaneously sorted the items according to their category relationship, whereas about 60% of the fourth graders did so.

The results of this study suggest important developmental changes in organizational behavior. Similar to the findings by Frankel and Rollins (1985) and Lange (1973), the younger subjects demonstrated less clustering during recall for stimuli containing weak interitem associativity. Obviously, associative relations play a substantial role in mediating organization in young but not in older elementary school children's memory. The findings

confirm the assumption that associative pairing may be the only organizational structure available to young children, whereas older children usually demonstrate greater flexibility (cf. Bjorklund, 1985; Bjorklund & de Marchena, 1984; Lange, 1978).

It should be noted, however, that there are also several differences between Frankel and Rollins' study and the present study. First of all, category relatedness played a more important role in German children's memory performance, as indicated by its effect on recall organization and recall. In addition, for German second graders, the presence of low associations negatively affected clustering but not recall, whereas low associativity negatively influenced recall but not clustering of the older children. In contrast, Frankel and Rollins' kindergartners demonstrated less clustering and recall of stimuli containing low interitem associations, whereas older subjects' recall organization and recall was not affected by low interitem associativity.

Interpreting these differences is not an easy task. They may be due to the fact that the two studies differ in several aspects: for example, Frankel and Rollins used kindergarten children rather than second graders as the youngest age group, and their design was multitrial compared to the single-trial memory task in the present study. However, the finding that fourth graders' performance across the two studies differed considerably cannot be accounted for simply by differences in design. It may be that these performance differences reflect cultural differences. Frankel and Rollins reported that their fourth graders displayed little organization at study, whereas the fourth graders in the present study sorted items taxonomically. This finding is consistent with the results of a cross-cultural study by Schneider, Borkowski, Kurtz, and Kerwin (in press) comparing American and German third graders' performance in a sort-recall task. In that study, American children displayed random input organization at pretest, while German third graders showed high levels of clustering. Interestingly, these significant differences in levels of input clustering disappeared after a short training procedure, indicating that the production deficiencies in American children could be easily eliminated. Schneider et al. attributed these differences to differences in school curriculum.

The distinction between sorting the items and studying the items made in the present study proved useful. Most subjects who sorted the items perfectly according to their category structure also used more effective study strategies like rehearsal and self-testing (5 of the 6 second graders and 21 of the 28 fourth graders did so). Most subjects who did not engage in sorting the items also failed to show any sign of a systematic study strategy (67% of the second graders and 38% of the fourth graders belonged to this category). Contrary to expectations, few children in either age group (seven second graders and four fourth graders) were inconsistent



in their sorting and studying behavior (e.g., very effective during sorting and ineffective during studying or vice versa). This finding may suggest that those children who spontaneously engaged in effective sorting behavior did have some knowledge of the value of the strategy and the relation between the strategy and the purpose of the task, that is, knowledge about the means-goals relationship (Paris, 1978).

The impact of the subjects' concepts about the value of different strategies for the memory task was more directly assessed by relating strategy use and memory performance to different components of metamemory. General metamemory was more closely (but nevertheless only moderately) correlated with second graders' strategy use and recall than task-related metamemory, which did not seem to have any impact on young children's memory behavior and performance. Moreover, both components of metamemory appeared to be independent for the younger subjects, and were also not significantly correlated with their judgments concerning the perceived efficacy of memory strategies. This indicates that relevant and stable preexisting domain-specific strategy knowledge was probably not available to most younger subjects. On the other hand, in the older subjects, the components of metamemory were related, and individual differences in task-related metamemory were important predictors of strategy use and recall, even after the influence of verbal and nonverbal intelligence had been partialled out. Further, task-related metamemory predicted judgments concerning the perceived efficacy of memory strategies. As can be seen from the scale values for paired comparison judgments in Table 3, fourth graders judged grouping and repetition strategies as more effective than the labeling and looking strategies. Interestingly, fourth graders in the present study evaluated the efficacy of different memory strategies very similarly to the second graders in Justice's (1986) study, whereas the second graders in the present study judged strategies very similarly to the kindergartners in Justice's study. This discrepancy may be due to differences in the stimulus materials used in the two studies, particularly the ease with which category relationships were detectable.

To summarize, converging evidence was found to support the claim that second graders are relatively unaware of the importance of categorization strategies for facilitating recall on sort/recall tasks, indicating that output organization (i.e., clustering during recall) is unvoluntarily guided by associations between items rather than by category grouping principles. Generally low and inconsistent metamemory judgments further indicate that deliberate strategy use is typically not found in this age group.

In contrast, fourth graders seem to be in a transitional state concerning the flexible and deliberate use of memory strategies. That is, approximately half of the sample of fourth graders showed systematic and strategic

behavior suited to facilitate recall, whereas the remaining subjects seemed unaware of the task requirements. Although the design of the experiment did not allow us to assess the functional relationship among task-specific metamemory, memory behavior, and performance, the close correlative relationship found among these variables indicates that voluntary strategies do affect recall of older children. However, it is important to note that the composition of sort/recall lists (i.e., the degree of interitem associativity) also affected memory behavior and recall of fourth graders with high metamemory scores. Apparently, these children are "transitional" concerning deliberate strategy use because they require somewhat structured materials to be able to capitalize on their metamemorial knowledge (otherwise they should be equally good on the low category relatedness/low interitem associativity list). Thus it appears that developmental shifts in memory behavior in sort/recall tasks are attributable to both developmental differences in semantic memory (Bjorklund, 1985) as well as to qualitative and quantitative differences in children's metamemory (cf. Schneider, 1985).

## REFERENCES

- Battig, W. F., & Montague, W. E. (1969). Category norms for verbal items in 56 categories: A replication and extension of the Connecticut category norms. *Journal of Experimental Psychology Monographs*, **80**, 1-46.
- Bjorklund, D. F. (1985). The role of conceptual knowledge in the development of organization in children's memory. In M. Pressley & C. J. Brainerd (Eds.), *Basic processes in memory development: progress in cognitive development research* (pp. 103-142). New York: Springer-Verlag.
- Bjorklund, D. F., & de Marchena, M. R. (1984). Developmental shifts in the basis of organization in memory: The role of associative versus categorical relatedness in children's free-recall. *Child Development*, **55**, 952-962.
- Borkowski, J. G., Peck, V. A., Reid, M. K., & Kurtz, B. (1983). Impulsivity and strategy transfer: Metamemory as mediator. *Child Development*, **54**, 459-473.
- Cattell, R. B., & Weiss, R. H. (1977). *Grundintelligenztest CFT 2 Skala 2*. Braunschweig: Westermann. (6. Aufl.).
- Cavanaugh, J. C., & Borkowski, J. G. (1980). Searching for metamemory-memory connections: A developmental study. *Developmental Psychology*, **16**, 441-453.
- Cavanaugh, J. C., & Perlmutter, M. (1982). Metamemory: A critical examination. *Child Development*, **53**, 11-28.
- Chechile, R. A., & Richman, C. L. (1982). The interaction of semantic memory with storage and retrieval processes. *Developmental Review*, **2**, 237-250.
- Corsale, K., & Ornstein, P. A. (1980). Developmental changes in children's use of semantic information in recall. *Journal of Experimental Child Psychology*, **30**, 231-245.
- Fippinger, F., & Rieder, O. (1978). *Allgemeiner Schulleistungstest AST 3*. Weinheim: Beltz.
- Flammer, A., Burkhardt, C., Jann, M., Reisbeck, C., & Stadler, S. (1985). Typikalitätsnormen für Exemplare von dreizehn Begriffen anhand der Publikation durch deutschschweizerische Studenten. *Sprache und Kognition*, **4**, 10-25.
- Frankel, M. T., & Rollins, H. A. (1982). Age related differences in clustering: A new approach. *Journal of Experimental Child Psychology*, **34**, 113-122.
- Frankel, M. T., & Rollins, H. A. (1985). Associative and categorical hypotheses of organization in free recall of adults and children. *Journal of Experimental Child Psychology*, **40**, 304-318.
- Justice, E. M. (1986). Developmental changes in judgments of relative strategy effectiveness. *British Journal of Developmental Psychology*, **4**, 75-81.

- Kee, D. W., & Bell, T. S. (1981). The development of organizational strategies in the storage and retrieval of categorical items in free recall learning. *Child Development*, **52**, 1163-1171.
- Kreutzer, M. A., Leonard, C., & Flavell, J. H. (1975). An interview study of children's knowledge about memory. *Monographs of the Society for Research in Child Development*, **40**, Serial No. 159.
- Lange, G. W. (1973). The development of conceptual and rote recall skills among school age children. *Journal of Experimental Child Psychology*, **15**, 394-407.
- Lange, G. W. (1978). Organization-related processes in children's recall. In P. A. Ornstein (Ed.), *Memory development in children* (pp. 101-128). Hillsdale, NJ: Erlbaum.
- Marshall, G. R., & Cofer, C. N. (1970). Single-word free-association norms for 328 responses from the Connecticut Cultural Norms for verbal items in categories. In L. Postman & G. Keppel (Eds.), *Norms of word associations* (pp. 321-360). New York: Academic Press.
- McNemar, Q. (1969). *Psychological Statistics (4th ed.)*. New York: Wiley.
- Moely, B. E. (1977). Organizational factors in the development of memory. In R. V. Kail, Jr. & J. W. Hagen (Eds.), *Perspectives on the development of memory and cognition* (pp. 203-236). Hillsdale, NJ: Erlbaum.
- Murphy, M. D. (1979). Measurement of category clustering in free recall. In C. R. Puff (Ed.), *Memory organization and structure* (pp. 51-83). New York: Academic Press.
- Naus, M. J., & Ornstein, P. A. (1983). Development of memory strategies: Analysis, questions and issues. In M. T. H. Chi (Ed.), *Trends in memory development research* (pp. 1-30). Basel: Karger.
- Nelson, K., & Brown, A. L. (1978). The semantic-episodic distinction in memory development. In P. A. Ornstein (Ed.), *Memory development in children* (pp. 233-241). Hillsdale, NJ: Erlbaum.
- Ornstein, P. A., & Corsale, K. (1979). Organizational factors in children's memory. In C. R. Puff (Ed.), *Organization, structure, and memory* (pp. 219-257). New York: Academic Press.
- Ornstein, P. A., & Naus, M. J. (1985). Effects of the knowledge base on children's memory strategies. In H. W. Reese (Ed.), *Advances in child development and behavior* (pp. 113-148). New York: Academic Press.
- Palermo, D., & Jenkins, J. J. (1964). *Word association norms*. Minneapolis: Univ. of Minnesota Press.
- Paris, S. G. (1978). Coordination of means and goals in the development of mnemonic skills. In P. A. Ornstein (Ed.), *Memory development in children* (pp. 259-273). Hillsdale, N.J.: Erlbaum.
- Pressley, M., Forrest-Pressley, D. L., Elliott-Faust, D., & Miller, G. (1985). Children's use of cognitive strategies, how to teach strategies, and what to do if they can't be taught. In M. Pressley & C. J. Brainerd (Eds.), *Cognitive learning and memory in children* (pp. 1-47). New York: Springer-Verlag.
- Roenker, D. L., Thompson, C. P., & Brown, S. (1971). Comparison of measures for the estimation of clustering in free recall. *Psychological Bulletin*, **76**, 45-48.
- Schneider, W. (1985). Developmental trends in the metamemory-memory behavior relationship: An integrative review. In D. L. Forrest-Pressley, G. E. MacKinnon, & T. G. Waller (Eds.), *Metacognition, cognition and human performance* (Vol. 1, pp. 57-109). New York: Academic Press.
- Schneider, W., Borkowski, J. G., Kurtz, B. E., & Kerwin, K. (in press). Metamemory and motivation: A comparison of strategy use and performance in German and American children. *Journal of Cross-Cultural Psychology*.
- Thurstone, L. L. (1927). A law of comparative judgment. *Psychological Review*, **34**, 237-286.