

CHAPTER 9

Situational Context Features And Early Memory Development: Insights From Replications Of Istomina's Experiment

Wolfgang Schneider and Marcus Hasselhorn

Although the need for replication of empirical studies has been emphasized by numerous researchers in various fields, the total number of true replication studies seems comparably small, at least in the case of experimental studies conducted in the social sciences. How can we account for this obvious discrepancy? In our view, several factors contribute to this unsatisfactory situation. First, conducting a true replication study does not seem very attractive to many because the benefit is small at least. If the replication study confirms the findings of the original study, this result gives further credit to the author of the original study. As the author of a replication study usually does not invest much of his or her creativity in the study but simply tries to follow the original design as closely as possible, the successful replication does not enhance the researcher's reputation other than illustrating that he or she is principally able to carry out experimental research properly. If a replication is unsuccessful, the case is even worse: Given the fact that in most cases it is impossible to exactly reconstruct the conditions of the original experiment, the author of the replication study may have a very difficult time convincing colleagues or reviewers that the deviating findings of the replication study do indeed point to flaws in the original study and cannot be explained by deviations in the experimental procedure of the replication study or by differences in the recruitment of subjects, and so on.

As a consequence, it appears that authors of replication studies should try to reconstruct the experimental conditions of the original study as closely as possible. In most cases, this may lead to additional problems because the detailed information needed for such a purpose is usually not provided in published materials. From our experience, getting access to more detailed

information means a considerable investment of both time and energy, going far beyond the amount of effort typically invested in experimental studies.

A third and even more important problem is that such an effort is usually not well-received by editors of scientific journals. It appears to us that demonstrations of replicability do not rank particularly high on journal editors' priority lists. Again, the situation seems even worse in the case of failure to replicate because this generally implies nonsignificant results, a finding regarded as uninteresting and of little scientific value by general journal policy (cf. later a recontinuation of this point).

Given these problems, it is not surprising that the number of published replication studies is that small. In our view, this is really unfortunate. As many experimental studies are based on relatively small sample sizes, it is by no means obvious that their findings represent stable and easily replicable effects. Even if facts seem established at a specific point in time, they may only have transient validity in a rapidly changing society.

It is the major purpose of this chapter to demonstrate how our understanding of young children's memory development can be improved by way of replication efforts of classical studies. The example chosen concerns the impact of situational context on young children's memory performance, a topic first systematically addressed in an important study conducted by Z. M. Istomina in 1948. During the past 40 years the results of this study have considerably influenced our view of the impact of motivational and social variables on preschool and kindergarten children's memory development. It is interesting to note that, at least to our knowledge, no replication study was conducted before the mid-1980s. In the remainder of this chapter, we want to address this issue in more detail. After a short description of Istomina's study, we first focus on the question of why there were no replications of Istomina's experiment for almost 40 years. Next, the results of two recent replication studies are discussed in detail, with an emphasis on the reasons provided by the authors for conducting these studies. Although the results of these two studies differ in several aspects, they suggest the conclusion that Istomina's findings cannot be replicated. However, although these two studies told us a lot about possible methodological problems of Istomina's study, findings of a more recent replication study indicate that this may not be the whole story. This replication study, consisting of two related experiments, will be dealt with in more detail in the final part of this chapter. Interestingly enough, the first of these experiments indeed replicated Istomina's findings, although it improved over the methodological weaknesses of the original study. A closer inspection of procedural details revealed that specific problems associated with the context manipulation and not explicitly considered in the former replication studies could have contributed to this outcome. This possibility was systematically explored in a second replication experiment which provided a more complete picture of potential problems with replicating the

Istomina experiment. The results of this study serve as a basis for the discussion of possible implications and for concluding comments on this topic.

ISTOMINA'S THEORETICAL APPROACH

Istomina was primarily interested in the question of when and under which circumstances voluntary, deliberate memorization first emerges. Her working hypothesis was based on Soviet theories about the development of voluntary memory in children (cf. Leontjev, 1928, cited by Leontjev, 1977). According to this theoretical approach, memory processes are not independent but integrated into some other activity during the preschool age period and are thus involuntary. Later in development, these processes are transformed into internal acts, thereby becoming purposeful and voluntary. Istomina assumed that the following preconditions are required for such a transformation: First, there should be motives and/or motivational incentives making mnemonic goals meaningful; second, these mnemonic goals should have some concrete reality; and third, the child should have means (strategies) available supporting voluntary memorization. The aims of Istomina's study were (a) to explore the conditions under which a young child is able to deliberately remember information, and (b) to determine the forms in which the acts of memorization first occur, with a focus on the memory operations by which they are affected.

DESIGN AND RESULTS OF THE ISTOMINA EXPERIMENT

Istomina created two settings—a game and a lesson—for children to remember words. Both conditions of the experiment were similar in form but differed functionally. In the lesson condition, the mnemonic goal was explicitly stated by the experimenter and was the only goal of the task. In the game condition, the children had to set goals by themselves (for details of the procedure, see Part III, Introduction). Istomina's hypothesis was that the game situation would be more familiar and also more motivating, and thus, young children would remember more in the game condition compared to the lesson condition.

Istomina tested a total of 60 children of the four ages 3–4, 4–5, 5–6, and 6–7. The same children participated in both conditions. As Istomina expected, recall improved notably with age, and children's recall in the game condition was superior to lesson recall at every age level. See Table 9.1 for an overview of these findings. Particularly poor memory was observed in the 3-year-olds.

Informal and qualitative observations of the children's behavior led Isto-

Table 9.1. Mean number of items recalled by preschoolers and kindergartners as a function of age and activity (Lesson vs. Game)*

Age (yr)	Activity	
	Lesson	Game
3-4	0.6	1
4-5	1.5	3
5-6	2.0	3.2
6-7	2.3	3.8

*Reconstructed from "The Development of Voluntary Memory in Preschool-age Children," by Z.M. Istomina, 1977, *Soviet developmental psychology* (M. Cole, Ed.), pp. 109, 111, New York: Sharpe. Reprinted by permission.

mina to conclude that the youngest age group did not understand the goal of remembering the items to buy. The major behavioral difference between 3- and 4-year-olds was that the latter group seemed to make an effort to remember. Regarding the two oldest age groups (5-6 and 6-7 years), Istomina assumed that these children already were aware of the mnemonic goal, and that they were capable of achieving that goal by selecting the strategy of rehearsal. However, although these children behaved similarly in the two experimental conditions, they performed less well in the lesson condition. It should be noted that no statistical tests of the data were provided.

ISTOMINA'S INTERPRETATION OF RESULTS

Istomina assumed that the consistent recall differences favoring the game over the lesson condition were due to differential motivational incentives for the children. Remembering presumably was an intrinsically important goal and had real meaning for the children in the game situation, especially in comparison to the lesson situation. Consequently, they probably showed greater interest in the task. In Istomina's view, differences in motivation could also account for the finding that even for the two oldest age groups better performance was obtained for the game condition as compared to the lesson condition, although attempts to deliberately memorize the items were similarly observed in both conditions. Her explanation was that

the first awareness of remembering and recall as an explicit goal occurs much more readily if the sense of such a goal follows naturally from the motivating factor intrinsic in the overall situation . . . rather than lying in a more complicated and indirect relationship between goal and incentive, as was the case in the laboratory test experiment. (1975, p. 40)

For the following 40 years, this study has been cited widely as evidence that (a) deliberate remembering is not evident before 4 or 5 years of age, and (b)

preschoolers remember more in the context of play activities carried out in familiar settings where memory is not an explicitly stated goal than they do in situations involving deliberate retention (cf. Weissberg & Paris, 1986).

REPLICATIONS OF ISTOMINA'S EXPERIMENT

Before focusing on those three studies that we conceive of as close replications of Istomina's study, we wish to speculate about possible reasons it took so long before replications of Istomina's experiment were carried out. One obvious reason is that Istomina's study was not known to researchers from Western countries until its English translation was published in 1975. In our view, another important reason is that nobody seriously questioned Istomina's results because they were intuitively appealing. Neisser (1982), in his commentary regarding Istomina's experiment, gives a probably representative characterization in that he found the study to be both charming and convincing. Reviews of Soviet investigations of memory and cognition provided particularly in the seventies by Michael Cole, John Meacham, and James Wertsch (e.g., Cole, 1978; Meacham, 1977; Wertsch, 1981) lead to an increasing popularity of Soviet conceptions of cognitive development among American and European researchers. In particular, the reception of Soviet theories of cognitive development has stimulated new interest in issues like the influence of context on development (e.g., Ceci, Bronfenbrenner, & Baker, 1988), and the importance of ecological validity for our proper understanding of memory phenomena (cf. Perlmutter, 1988). Undoubtedly, the focus on issues like environmental context and ecological validity apparent in many recent studies on memory development has increased our knowledge about memory phenomena considerably (cf. for a review Schneider & Pressley, 1989). Accordingly, there seemed to be no reason to question Istomina's finding that young children's memory processes are influenced by contextual factors.

Given that Istomina's findings square very well with contemporary views of memory development, it seems surprising that replication studies were carried out at all. It is indeed difficult to explain why two research groups in the United States and Germany (Weissberg & Paris, 1986; Schneider & Brun, 1987) independently decided to replicate Istomina's study at about the same time. It was probably more or less by accident that both groups discovered possible methodological problems with Istomina's experiment that seemed serious enough to warrant a replication study.

Before discussing these problems in more detail, we want to make one important point clear: The goal of all replication studies reviewed in the remainder of this chapter was to explore the value of Istomina's findings for current memory research. In our view, it is impossible to replicate Istomina's

experiment in a strict sense, mainly because possible influences of cohort and cultural setting cannot be dealt with adequately in contemporary replication studies. For example, running an errand to a grocery store in order to buy food and other things for the kindergarten group probably had a different meaning for the Soviet children in 1948, than for American and German youngsters of the mid-1980s. Given that access to food was restricted shortly after World War Two (note that children were given "admission slips"), it appears that the shopping task was of great personal value and probably resulted in a feeling of responsibility in the Soviet children that could not be equally induced in American and German children participating in the replication experiments (cf. Folds-Bennett, chapter 10). Thus we should keep in mind that the outcomes of the various replication studies do not tell us much about whether Istomina was right or wrong; they only give information about the value of Istomina's findings for current memory research.

METHODOLOGICAL PROBLEMS WITH ISTOMINA'S STUDY

Both Weissberg and Paris (1986) and Schneider and Brun (1987) emphasized that Istomina's study contains several methodological problems. Although Istomina provided a very detailed description of her study, including a number of highly interesting analyses of individual cases, it is very difficult to reconstruct her study exactly because the procedures were specified vaguely. Weissberg and Paris (1986) inferred that the order of activities was not counterbalanced. This is probably right, since we know from the literature that the need for counterbalancing the order of activities was typically not realized in Soviet studies of memory development conducted in the 1940s and 1950s (cf. Schneider & Pressley, 1989, for a review). However, this cannot be decided on the basis of Istomina's description of the procedure, because the only information about this point is that the two experiments were carried out "simultaneously."

Another methodological problem was that different word lists were used in the two conditions, and that the word lists varied in unspecified ways between subjects. Further, a closer look at the reports of individual subjects clearly shows that the procedure was not strictly standardized. The experimenter occasionally repeated the names of the objects in the game condition but did not do so in the lesson condition. Thus one is inclined to assume that the generally superior recall found for the game condition can at least partly be explained by methodological flaws systematically biased against the lesson condition.

The results from a few experiments not closely replicating Istomina's study seemed to confirm this suspicion. By using a performance prediction paradigm, Wippich (1981) found that the game condition did not yield superior

recall in preschool children, as compared to the lesson condition. In a more recent investigation also based on the performance prediction paradigm, Arbinger and Kubsda (1987) basically replicated Wippich's findings. Further, Wippich (1985) contrasted laboratory with game conditions for various memory paradigms. In general, older children always performed better than younger children in these studies, but most critically, memory context never affected performance. Given these findings, it seemed to make sense to carry out close replications of Istomina's study, mainly because of the influence it has had on current claims concerning young children's memory.

REPLICATION STUDY 1: THE WEISSBERG AND PARIS (1986) EXPERIMENT

The primary goal of this study was to replicate Istomina's experiment and to correct the methodological shortcomings mentioned above. That is, the order of activities was counterbalanced, the same words were used for all children, and lists of items were counterbalanced across game and lesson conditions. A total of 96 children participated in the study. There were four age groups, each with 12 boys and 12 girls. Subjects ranged in age from 3 to 7 years.

Weissberg and Paris' (1986) study differed from the original experiment in that two different game settings and two lesson activities were used. Word lists were presented either as part of shopping or party games or as part of corresponding deliberate lessons. Thus, each child participated in four memory tasks—remembering food items or names as part of a game or a lesson activity—in a 4(age) x 2(sex) x 2(activity) x 2(mnemonic object) design. The experiment was well-controlled: The presentation orders of the four situations were counterbalanced within age groups, and the sequences of mnemonic objects were counterbalanced within activities for each group.

The procedure concerning the shopping situation was modeled closely after Istomina's description (see above, and Chapter 10). However, the scenario differed in several aspects. First, the social component of the original study was missing. That is, while the game was played in two adjoining rooms (a "kindergarten" and a "store") in the Istomina study, children in the Weissberg and Paris study were tested individually in one empty room. Children were instructed to buy a list of items from the play grocery store, but were not told that they should buy items for the kindergarten (as were the subjects in Istomina's study). Second, Paris and Weissberg used a list of six items (as compared to the five items used by Istomina), which was read twice in the case of Weissberg and Paris and once in Istomina's experiment.

In the "party" situation, children were shown seven stuffed animals in a circle. Children were asked to pick the animal they liked the most, which was usually given the subject's first name (e.g., Johnny). Children were then told

that the six remaining animals were having a party, and that Johnny would also like to be at that party. Children were then told that Johnny's problem was that he did not know the names of the other animals. They were then given the names of the animals at the party twice, and were asked to introduce the six animals to Johnny after 60–90 seconds had elapsed.

The lesson counterparts to both games required subjects to recall the same words after equivalent delays for the sole purpose of remembering. Instructions were closely modeled after those of Istomina's study.

As a major result concerning the 2-item lists, Weissberg and Paris reported that children recalled more food items than animal names. In the following, we focus on the food items which represent the replication part of the Weissberg and Paris study. Data from the replication study are shown in Table 9.2. As Weissberg and Paris pointed out, the different patterns of results are apparent immediately. While children in Istomina's experiment recalled more items in the game condition than in the lesson condition, the opposite was true in Weissberg and Paris' study in that recall turned out to be consistently better in the lesson condition. The sex-by-activity interaction was significant in the Weissberg and Paris experiment, indicating that while both boys and girls recalled more in the lesson activities than in the game, boys remembered more than girls in game activities and less in lesson activities.

A surprising finding of Weissberg and Paris' study was the large number of subjects who showed signs of rehearsal. Regarding the food items, more children (about 63%) rehearsed in the lesson condition than in the game condition (24%). Although older children rehearsed more often than the younger subjects, it seems remarkable that nearly half of the 3-year-olds rehearsed words spontaneously in the lesson activity. Weissberg and Paris concluded from their findings that children's recall does vary in different contexts, but that the underlying reasons may reflect task-relevant knowl-

Table 9.2. Comparison of the Weinberg and Paris data with Istomina's data: Mean numbers of words recalled by age and activity

Age	<i>Istomina's Data</i>			Sex	<i>Food</i>		
	<i>Lesson</i>	<i>Play</i>	<i>Ratio*</i>		<i>Lesson</i>	<i>Play</i>	<i>Ratio</i>
3–4	.6	1.0	1.7	Male	3.0	2.4	.8
				Female	2.5	2.3	.9
4–5	1.5	3.0	2.0	Male	3.4	2.8	.8
				Female	3.5	2.7	.8
5–6	2.0	3.2	1.6	Male	3.9	3.5	.9
				Female	4.3	2.8	.7
6–7	2.3	3.8	1.7	Male	4.4	4.1	1.0
				Female	3.8	3.8	1.0
Mean total	1.6	2.8	1.8		3.7	3.1	.9

*The ratio is calculated by dividing the words recalled in play by the number recalled in the lesson condition.

edge and memory skills elicited in appropriate settings rather than global motivational dispositions.

REPLICATION STUDY 2: THE SCHNEIDER AND BRUN (1987) EXPERIMENTS

As noted earlier, Schneider and Brun shared Weissberg and Paris' view that methodological problems could be responsible for the context effect found in Istomina's study. In order to clarify this point, Schneider and Brun first tried to contact Z. M. Istomina directly at the Pedagogical Institute of Moscow University. Unfortunately, there was no reply. Consequently, Schneider and Brun decided to carry out research to replicate Istomina's experiment and to improve on her experimental methods. Although Weissberg and Paris were convinced that there was no counterbalancing in Istomina's study, Schneider and Brun were not equally confident, mainly because Istomina's description of the experimental design was so vague.

Thus they decided to first carry out an experiment which was aimed at reconstructing the experimental conditions presumably used by Istomina including the methodological shortcomings as they perceived them. The expectation was that it would be possible to replicate Istomina's findings under these circumstances. The second experiment was designed to correct shortcomings of the original procedure. That is, the order of play and lesson conditions was counterbalanced, lists of items were counterbalanced across the two experimental conditions, and precautions were taken to provide identical, standardized instructional settings in the two experimental conditions. Accordingly, the second experiment was thought to represent a more appropriate test of the context hypothesis.

Experiment 1. A total of 60 children participated in the experiment. Children were recruited from four kindergartens in the Munich area. There were only two age groups in this experiment, with 15 boys and 15 girls in each. The mean ages of the younger and older groups were about 4 years and 6 years, respectively.

The procedure was closely modeled after that used by Istomina and described above. However, there were deviations from Istomina's design similar to those already noted for the Weissberg and Paris study. That is, the social ("Play kindergarten") component was missing; all children were tested individually in one single room. Further, the two word lists included 8 items each, as compared to the 5-item word lists used by Istomina. Eight items were given in order to avoid ceiling effects; a pilot study with 6-year-olds had shown that shorter lists do not yield much variance in this age group.

In short, the results revealed that Schneider and Brun were able to replicate

Istomina's findings. On average, 6-year-olds recalled more than 4-year-olds, and memory performance in the game condition was significantly better than performance in the lesson condition for both age groups. No further main effects or interactions were found. The liberal instructional setting, that is, the possibility of returning to the experimenter and asking her to repeat the items, appeared to have different impact in the two experimental conditions: While in the lesson condition children did not find it necessary to ask for a repetition, the latter was more common in the game condition. Not surprisingly, those who asked the experimenter to repeat the items also recalled more than those who listened to the stimulus list only once. When initial recall (i.e., prior to any repetition by the experimenter) was used as the dependent variable, the effect of experimental condition was negligible. Schneider and Brun concluded from Experiment 1 that a major weakness of Istomina's experiment concerned the instructional setting: Better recall in the game condition may have been related to additional repetitions of the stimulus list in that condition.

Experiment 2. The only difference between Experiment 1 and Experiment 2 concerned the instructions given to the children. In both conditions of Experiment 2, children were told that they had to remember a list of items which the experimenter would read to them only once. In addition to the instruction given in Experiment 1, the experimenter explicitly informed the subjects that the word lists would not be repeated. Again, a total of 60 children was recruited. Half were about 4 years old, and half of them about 6 years of age. There were equal numbers of boys and girls in each group.

The results showed that only the main effect of age was significant. On average, 6-year-olds recalled more items than 4-year-olds. Although performance tended to be better in the game condition, there was no significant effect of activity. A significant interaction between order of activities and number of items recalled in the two experimental conditions indicated that, regardless of age, better recall was obtained for the experimental task that was presented first.

The data from Experiments 1 and 2 as well as Istomina's data are depicted in Table 9.3. It seems obvious that the strictly standardized instruction used in Experiment 2 affected memory performance particularly in the game condition. As a consequence, the game recall/lesson recall ratios in Experiment 2 were substantially lower than those obtained in Experiment 1 and in Istomina's study. Schneider and Brun (1987) concluded from this that Istomina's findings could not be replicated when stricter experimental controls were introduced. Thus the hypothesis that Istomina's findings were due in part to methodological shortcomings (which was already supported by the outcome of the Weissberg and Paris experiment) was also confirmed by the results of Schneider and Brun's second experiment.

Table 9.3. Comparison of the Schneider and Brun data with Istomina's data: Mean number of words recalled by age and activity

Age	<i>Istomina's data</i>			<i>Experiment 1</i>			<i>Experiment 2</i>		
	<i>Lesson</i>	<i>Game</i>	<i>Ratio</i>	<i>Lesson</i>	<i>Game</i>	<i>Ratio</i>	<i>Lesson</i>	<i>Game</i>	<i>Ratio</i>
4	1.05	2.00	1.8	1.63	3.56	2.2	1.59	2.08	1.3
6	2.15	3.50	1.7	3.06	5.46	1.8	2.53	2.79	1.1

*Data were collapsed across the 3-4- and 5-year-olds on the one hand and the 5-6- and 6-7-year-olds on the other hand.

A COMPARISON OF THE WEISSBERG AND PARIS (1986) AND SCHNEIDER AND BRUN (1987) EXPERIMENTS

At first glance, the experiments carried out by Weissberg and Paris (1986) and Schneider and Brun (1987) seem to convey the same message, namely, that Istomina's findings obtained in the 1940s are obscured by methodological flaws and at least cannot be generalized for children raised in the 1980s. However, a closer comparison of the two replication studies reveals that their results differed in several aspects. First, Weissberg and Paris found that their subjects generally recalled more items in the lesson condition, as compared to the game condition, whereas no significant difference between conditions was obtained by Schneider and Brun when stricter experimental controls were introduced in Experiment 2. A second difference concerned the frequency of rehearsal. In the Weissberg and Paris study, a surprising finding was the large number of children who rehearsed. On the other hand, only a few children in the Schneider and Brun study showed spontaneous verbal rehearsal. Possibly related to this, the children in the Weissberg and Paris study recalled considerably more than the children in the Schneider and Brun study, regardless of age and condition. This finding seems particularly surprising given that more items per set were used in the Schneider and Brun study.

How can we account for these differences? We can only speculate about the underlying causes. One possibility is that the repeated presentation of the word lists in the Weissberg and Paris study may have contributed to the high frequency of rehearsal observed in their sample. As those children who rehearsed words recalled significantly more words than those who did not rehearse, regardless of age, the overall higher recall observed in the Weissberg and Paris study could be due to the comparably higher incidence of rehearsal. If this holds true, the question remains why Weissberg and Paris's subjects rehearsed more in the lesson condition than they did in the game condition. In Weissberg and Paris's view, this finding can be attributed to the fact that there were more distractions involved in the game condition. Although this explanation seems to make sense, it does not account for

Schneider and Brun's finding that their subjects tended to recall more in the game condition than in the lesson condition.

Taken together, a systematic comparison of the two replication studies reveals a number of differences in findings that are very difficult to explain given the information at hand. While in both studies an attempt was made to replicate Istomina's design, a closer inspection of the procedural details shows that not only did the procedures and materials used in the two replication studies differ from each other but they also differed in several aspects from the original study. As noted above, the social component inherent in Istomina's study (i.e., Playing Kindergarten) was not included in the replications. Although we do not think that the failure to replicate Istomina's results is due to this aspect—in fact it could have increased the number of distractions involved in the game condition—it is certainly unfortunate that the implications of this aspect were not assessed. We are thus left with a dilemma: Although the results from the two replication studies indicate that several methodological problems could have biased Istomina's findings, the extent to which deviations from the original procedures and materials affected the outcomes of the replication studies is impossible to assess.

Given these problems, we decided to conduct another replication study just to explore the significance of other variables that had not been systematically controlled in the two previous replications.

REPLICATION STUDY 3: HASSELHORN AND SCHNEIDER (1990)

A closer inspection of Schneider and Brun's (1987) data reveals that reproduction probabilities for single items varied considerably in both experiments. We were led to the conclusion that this extreme variability could be due to effects of attractiveness or salience of the item materials. Apparently, children remembered the more attractive objects much better. We decided to carry out another replication experiment to explore more systematically the effects of attractiveness of item materials on recall. The question of specific interest was whether Schneider and Brun's (1987; Experiment 2) results (i.e., the failure to replicate Istomina's findings) could be replicated for both attractive and unattractive stimulus materials.

Experiment 1. To assess the attractiveness of item materials, a pilot study was carried out with 40 4- and 6-year-old children, with 20 subjects in each age group. Children were presented with a total of 40 photographs showing various objects. According to our judgment, 20 of these objects were probably attractive for most subjects, whereas the remaining objects were perceived of as rather unattractive. Children were given four sets of 10 items. Each set included five items rated as attractive and five items rated as

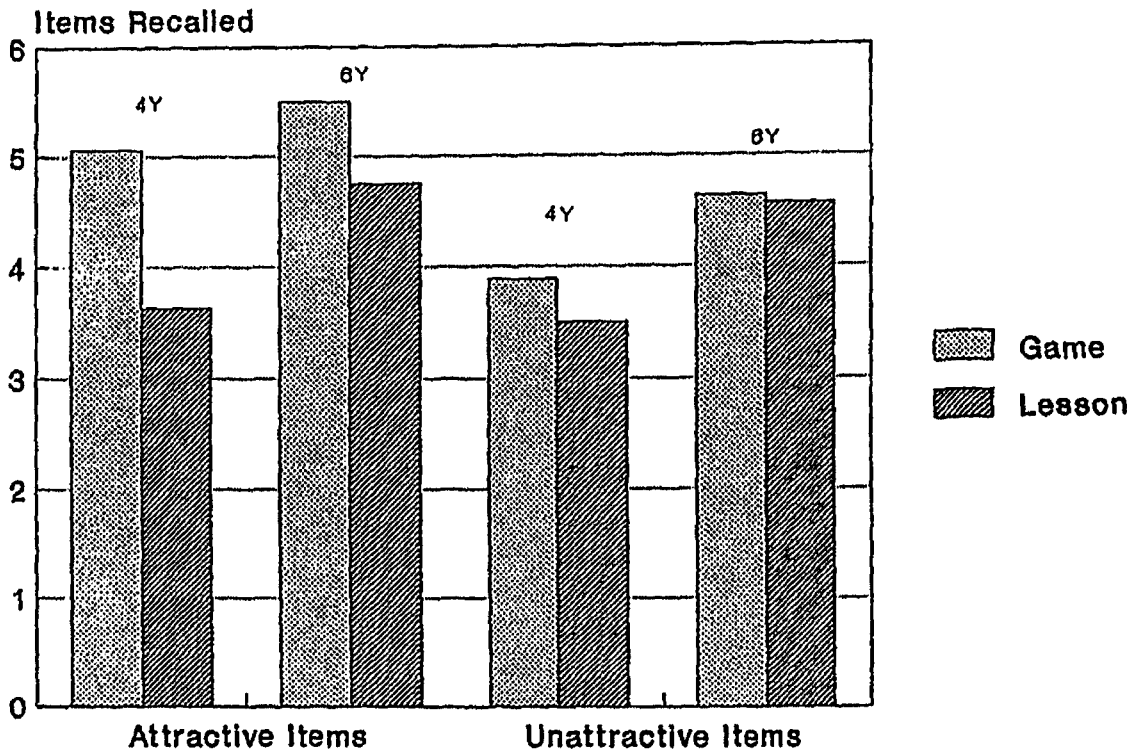
unattractive. The subjects were told to first choose those three items (out of 10) that they considered most attractive ("show me the three pictures that you like most"), and then to select those three items out of the remaining seven that they found least attractive ("show me the three pictures that you like least"). Attractiveness judgments were analyzed separately for the two age groups. The list of unattractive items chosen for Experiment 1 consisted of 8 items that were similarly classified as unattractive by 4- and 6-year-olds. The following items were considered unattractive by most children: toothpaste, vase, handkerchiefs, rock, screw driver, bulb, sponge, and cigarettes (Zahnpasta, Vase, Taschentuch, Stein, Schraubenzieher, Glühbirne, Schwamm, Zigaretten).

The list of attractive items chosen for Experiment 1 included 8 items selected as attractive by the majority of children in both age groups. The following objects were rated as attractive: Toy car, crayons, balloon, teddy bear, chocolate egg, engine, licorice, and sailboat (Spielzeugauto, Malstift, Luftballon, Teddybär, Schokoladenei, Eisenbahn, Lakritze, Segelboot).

A total of 128 children participated in Experiment 1. There were two age groups of children, with 64 children in each. The mean ages of the younger and older groups were about 4 and 6 years, respectively. While there were equal numbers of boys and girls in the younger group, the older group consisted of 24 boys and 40 girls. Children were recruited from kindergartens in the Göttingen West Germany area. With each age group, subjects were randomly assigned to the two activity (game vs. lesson) and attractiveness conditions. The procedure of this experiment differed from that used by Schneider and Brun in that items were not only read to the children but also presented visually at the same time. Another difference concerned the fact that items were given twice in this experiment but only once in the Schneider and Brun study.

Preliminary analyses revealed that there were no effects of sex. Thus data were collapsed across the sex variable. Figure 9.1 shows the mean number of items recalled as a function of age, item attractiveness, and activity. Analyses of variance revealed a significant effect of age ($F(1,120) = 22.2, p < .001$). Six-year-olds recalled more items than 4-year-olds. There were also main effects of attractiveness ($F(1,120) = 11.0, p < .001$) and activity ($F(1,120) = 13.4, p < .001$). However, these effects were qualified by a significant activity by attractiveness interaction ($F(1,120) = 6.0, p < .05$). Post-hoc comparisons revealed that better recall in the game condition was restricted to the attractive item materials, regardless of age.

In addition to children's recall, signs of rehearsal were also recorded as a simple dichotomous variable. Children who either showed lip movements or verbally repeated the words were classified as active rehearsers. Based on these selection criteria, about 47 percent of the older children and about 41 percent of the younger children were classified as active rehearsers. There



Note: This experiment used items per set

Figure 9.1. Mean number of items recalled as a function of age, item attractiveness, and activity (data from Hasselhorn & Schneider, 1990; Experiment 1).

were slightly more boys than girls among active rehearsers (50% vs. 38%), and less children rehearsed in the lesson condition than in the game condition. However, these differences did not prove statistically significant. The only effect approaching significance was observed for item attractiveness ($F(1,120) = 3.2, p < .10$): Although about 52% of the children presented with attractive items actively rehearsed the stimuli, only 36 percent of the children presented with unattractive items did so. Point-biserial correlations between rehearsal and recall were nonsignificant for both younger and older children ($r's = .04$ and $-.06$, respectively), indicating that the systematic relationship between rehearsal and recall reported by Weissberg and Paris was not found in this experiment.

Taken together, the results of this experiment seem to indicate that the context effect in favor of the game condition reported by Istomina can be replicated even after methodological problems of Istomina's study have been removed. However, the effect may not be as general as originally assumed, that is, restricted to learning materials that seem particularly attractive for young children. Thus the results of Experiment 1 suggest that it does not make much sense to talk about general effects of *the* context because different features of the situational context may turn out to be relevant in memory studies with young children. It seems that not only *objective* environmental features of the learning situation (game vs. lesson) but also *subjective*

tive features like the perceived attractiveness of stimulus materials contribute to this effect. Apparently, different context features interact in influencing young children's memory performance.

If this is so, a reasonable next step seems to explore other features of situational context possibly relevant for memory performance. In order to find out about additional context features, we decided to interview our experimenters of Experiment 1, asking them to reconstruct the procedural details as closely as possible. Further, we analyzed some tape recordings of experimental sessions which originally were produced as instructional materials in developmental psychology classes at the University of Göttingen. Both the discussions with the experimenters and the inspections of tape recordings confirmed our suspicion that the game and lesson conditions differed in several ways. In addition to the environmental context features, the experimenters' behavior also varied across activities. It appeared that the game condition seduced the experimenters to communicate in a more friendly, motivating, and empathic way with the children than they did in the more sterile lesson condition.

There were several signs supporting the assumption that the type of activity influenced the experimenters' behavior. First, it seemed that the experimenters established a more *personal contact* with the subjects in the game condition, as compared to the lesson condition. For example, in the lesson condition the experimenter would start with instruction such as "Let's play a game together" and then explain the procedural details to the child. In the game condition, the experimenter would typically begin with "Let's play doing an errand in a grocery store. You probably have done that before with your mother, haven't you?" In most cases, this question led to a short conversation between child and experimenter, with the child informing the experimenter about his or her experiences in a shopping situation. Such conversations did not take place in the lesson condition.

Further, the tape recordings revealed that the pitch of the experimenters' voices differed between the two conditions. That is, the experimenter's voice tended to be lower and more monotone in the lesson condition, as compared to the game condition. Moreover, there seemed to be systematic differences regarding the way the experimenter verbally reinforced children's responses in the two conditions. These differences were not quantitative but qualitative in nature. For example, in the lesson condition, the experimenter's usual comment on correct responses was "right" (*ja*), as compared to a "great" (*sehr gut*) following a correct response in the game condition. Finally, our interview with the experimenters revealed that they tended to nonverbally encourage children to persist and to try recall of items (e.g., via gestures) in the game condition but not in the lesson condition.

Experiment 2. Given these specific problems, we decided to conduct another replication experiment. The major goal of this experiment was to

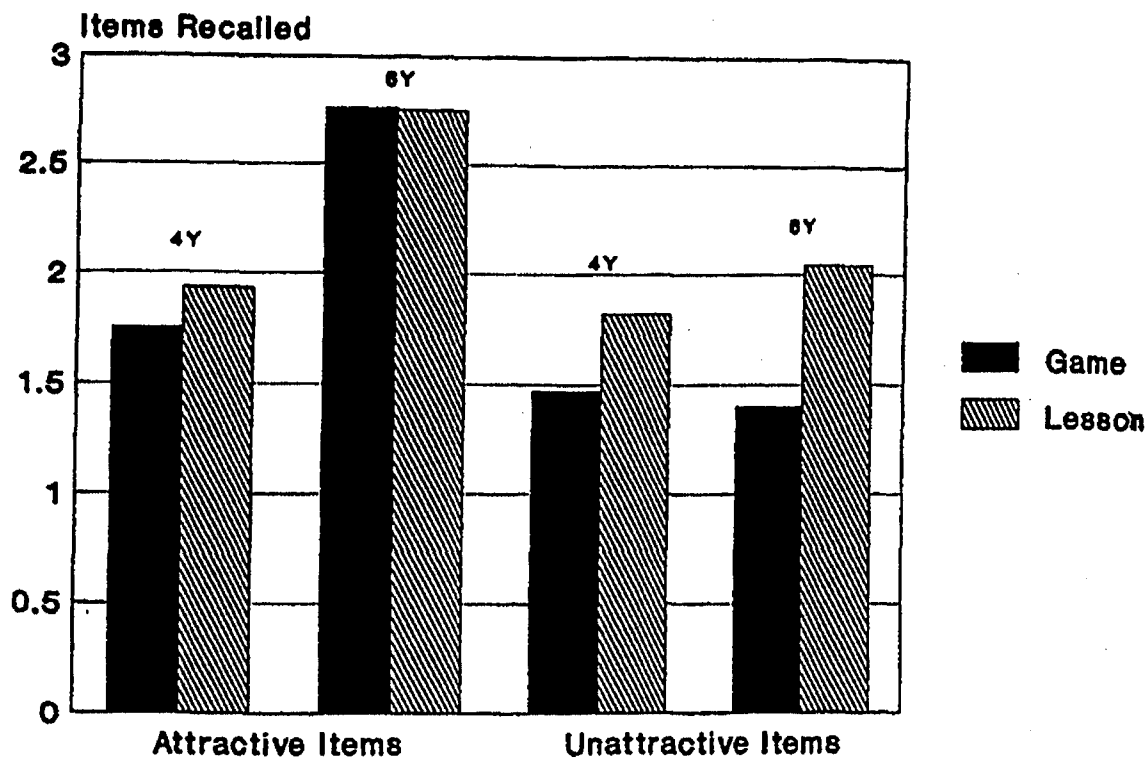
explore whether the context effect found in Experiment 1 with attractive stimulus materials could be replicated with experimenters being informed about the possible influences of environmental context on their behavior in the experimental session. A second aim of Experiment 2 was to explore the impact of item presentation mode (visual vs. acoustic) on memory performance.

Experiment 2 differed from Experiment 1 in two ways. First, the experiments were informed about the influence of the game situation on their behavior and were instructed to behave similarly in the two activity conditions. Second, item attractiveness was varied *intraindividually* and was based on individual attractiveness judgments.

A total of 112 4- and 6-year-old children participated in Experiment 2, with 56 subjects in each age group. There were the same number of boys and girls in each group. Subjects were recruited from 5 kindergartens in the Göttingen and Munich areas. Children were presented with a list of 10 items (5 attractive and 5 unattractive items). In the visual presentation mode condition, children were shown the photographs of the to-be-recalled objects, whereas items were read to the children in the acoustic presentation mode condition.

Again, preliminary analyses revealed no significant effects of sex on recall. Thus, recall data were collapsed across this variable and analyzed separately for attractive and unattractive items (cf. Figure 9.2). Accordingly, a two (age groups) by two (presentation modes) by two (activities) by two (attractiveness levels) design was used, with the last factor repeated within subjects. With the exception of presentation mode, all main effects were significant. Analyses of variance revealed a main effect of activity ($F(1,104) = 7.7, p < .01$), indicating that children recalled more in the lesson condition than in the game condition. The main effects of age ($F(1,104) = 19.3, p < .001$) and item attractiveness ($F(1,104) = 26.8, p < .001$) were qualified by a significant age by attractiveness interaction ($F(1,104) = 12.3, p < .01$). Post-hoc comparisons revealed that the age effect was restricted to attractive items, and that the attractiveness effect held only for the 6-year-olds. Undoubtedly, the most interesting finding of Experiment 2 concerned the *reversed* context effect: It was not in the game condition but in the lesson condition that children recalled more objects, regardless of age. However, this context effect was qualified by a significant three-way interaction between activity, age, and presentation mode ($F(1,104) = 6.0, p < .05$), indicating that older children's better recall in the lesson setting was restricted to the unattractive items. Thus the results of Experiment 2 replicated those obtained by Weissberg and Paris (1986).

As in Experiment 1 children were classified as active rehearsers versus nonrehearsers. While about 48% of the total sample showed signs of rehearsal, a two (age groups) by two (presentation modes) by two (activities) by two (sexes) analysis of variance only revealed a marginal age by sex interaction, indicating that there were more boys among the older rehearsers.



Note: This experiment used 5 items per set

Figure 9.2. Mean number of items recalled as a function of age, item attractiveness, and activity (data from Hasselhorn & Schneider, 1990; Experiment 2).

How can we account for the findings concerning the recall data? As noted earlier, one explanation offered by Weissberg and Paris is that the mnemonic goal of the task can be obscured by several task-irrelevant aspects of the game condition. Our observations of children's behavior in the game condition support such an assumption. For example it appeared that several children did not pay as much attention to the items read or shown to them than to details of the play grocery store ("What kind of stuff do you have in the grocery store?"). Similarly, signs of distractions became also apparent in the retrieval phase of the game condition ("I think some of the things I should buy were not at the grocery store"). Given the limited attentional capacity of these young children, it does not seem surprising that their recall was consistently low in the game condition.

How can we account for the fact that a (restricted) context effect was found in Experiment 1? We think that the problems related to the confounding noted above produced the findings to a considerable extent. Apparently, the experimenters' extremely friendly, motivating, and supportive behavior in the game condition helped to establish the mnemonic goal and more than compensated for the distractions, at least in the case of the attractive items. On the other hand, the findings of Experiment 2 indicate that the more neutral behavior shown by the experimenters had a different effect: Children seemed to distribute their cognitive resources across several goals and activ-

ities, with only one of them being the mnemonic goal implemented by the experimenter. In comparison, children have no problem using most of their cognitive capacity to recall items in the lesson condition, mainly because no secondary goals are available in this situation. Thus, we assume that the information processing capacity interpretation suggested by Weissberg and Paris is the most parsimonious in explaining the reversed context effects found in Experiment 2.

Alternatively, one could argue that it is not the *difference between* conditions but the *structure within* each of them that either gives or not the expective sample differences. So, by way of altering instruction style and triggering individual attractiveness to given objects in children, one can make the lesson condition more effective than the game condition. Likewise, by changing the friendliness of the game partner and giving different objects to remember, one can make the game condition ineffective. All these experimental variations concern the second structure, making the comparisons between settings quite unpredictable from study to study.¹

GENERAL DISCUSSION

At the very beginning of this chapter we discussed several problems replication studies have to deal with. We noted that experimenters failing to replicate a classic study almost certainly will experience difficulties with convincing the scientific community that the old study was indeed wrong. We think that the description of the three replications of Istomina's study experiment given in this chapter is well suited to illustrate the various problems that can be experienced in such a case.

To summarize, it appeared that the first two replication studies (Weissberg & Paris, 1986; Schneider & Brun, 1987) built a strong case against Istomina's original findings in that both studies were not able to replicate Istomina's results. This was all the more impressive as both studies improved Istomina's methodology and were conducted in different cultural contexts. However, a closer look at the outcomes of the two replication studies revealed that their findings differed in several aspects. Probably most important, Weissberg and Paris found a reversed context effect, indicating that American youngsters performed better in the lesson condition than they did in the game condition, whereas Schneider and Brun's German children still tended to be better (although not significantly) in the game condition than in the lesson condition after strict methodological controls were established. Another major difference between the two replications was that most children in the Weissberg and Paris study used verbal rehearsal, particularly in the lesson condition,

¹This is the interpretation preferred by Jaan Valsiner. We are grateful for his suggestion.

which helped in improving recall, whereas Schneider and Brun's subjects rarely used this mnemonic strategy.

The third replication study (Hasselhorn & Schneider, 1990) was carried out in order to explore the importance of additional context features not systematically considered in the two previous replications. The first experiment investigated the relevance of item attractiveness for memory performance, yielding the interesting as well as puzzling finding that Istomina's results could be replicated in the case of particularly attractive item materials. However, a closer inspection of procedural details revealed that not physical content per se, but different behavior of the experimenter in the two activity conditions seems to be responsible for the results. This suspicion was confirmed in a subsequent replication study that took care of the experimenter problem: In this experiment, the reversed context effect of Weissberg and Paris could be replicated, whereas the effect of item attractiveness was found to be restricted to the older children, positively affecting performance in both activity conditions.

What did we learn from all these replications?

First of all, it seems that even minimal differences in the experimental design and/or procedure can have dramatic effects on experimental outcomes. Our description of the replication studies revealed that they were not replications *in the strict sense*, but *conceptual* replications (see Finifter, 1975, for a more detailed discussion of this terminological distinction). That is, the major goal of the studies was to replicate Istomina's findings but not to reconstruct the design of the original study in detail. For example, we already referred to the fact that the social component ("Playing kindergarten") of Istomina's experiment was not included in the replications. This is also true for Experiment 1 of Schneider and Brun (1987), which came closest to a replication in the strict sense.

Why has no one tried to carry out a replication in the strict sense? There are several reasons. First, the information concerning procedural details given in Istomina's study are too vague to allow for a replication which matches every detail of the original study. Schneider and Brun's attempts to get more detailed information from Istomina were unsuccessful. Thus it appears that the ideal replication in the strict sense is impossible in the case of Istomina's study. But even if it were possible to exactly reconstruct Istomina's design, the inclusion of the social component would probably cause considerable problems. That is, the inclusion of a "Playing Kindergarten" setting would almost certainly add to the methodological problems because it is probably impossible to carry out a standardized experimental procedure using such a setting. Thus our conclusion is that a replication study in the strict sense is neither possible nor does it make much sense in the case of Istomina's experiment. We think that there is sufficient empirical evidence demonstrating the impact of methodological flaws on the outcomes of Istomina's study. Therefore, it

seems much more important to explore the possibility of conceptual replications.

By and large, the conceptual replications reviewed in this chapter suggest the conclusion that Istomina's findings cannot be generalized to current memory research. That is, those replication experiments which methodologically improved over Istomina's experimental design did not find any evidence for the context effect reported by Istomina. The context effect could only be demonstrated in those two experiments that either intentionally included the methodological flaws observed in Istomina's study (Schneider & Brun, Experiment 1) or that were shown to (unintentionally) include additional methodological problems at least from a strict experimental point of view (Hasselhorn & Schneider, 1990, Experiment 1). The results of these two experiments suggest that the induction of particularly motivating experimental conditions in the game situation (e.g., a strong personal commitment of the experimenters) was responsible for the effect in favor of the game context. These findings are due to the fact that the experimenters behaved differently in the two conditions, and must be interpreted as artifacts from a methodological point of view.

As indicated by the results of the Weissberg and Paris (1986) and Hasselhorn and Schneider (1990, Experiment 2) replications, it is more reasonable to expect a reversed context effect in strictly controlled experiments conducted in the eighties. It seems that there are at least two possible explanations for the fact that better memory can be observed in the lesson condition, as compared to the game condition. First, there is no doubt that the educational environments available to American and German children in the 1980s clearly differ from those typically experienced by Soviet children in the 1940s. Probably due to effects of preschool education, television, and parental influences, even 3- to 4-year-olds are capable of voluntary memory (cf. Baker-Ward, Ornstein, & Holden, 1984; Sodian, Schneider, & Perlmutter, 1986). American and German preschoolers are no longer bewildered by requests to remember lists of words. As indicated by the findings of Weissberg and Paris, they may even know how to use mnemonic strategies efficiently at a very early age.

The second explanation offered by Weissberg and Paris concerns the negative impact of task-irrelevant aspects of the game. That is, several distractors embedded in the game situation (e.g., attractive objects in the grocery store) may obscure the mnemonic goal of the task and instead generate new goals (like playing with those objects) that are detrimental to—if not incompatible with—the mnemonic goal implemented by the experimenter. Our observations (cf. Hasselhorn & Schneider, 1990, Experiment 2) do confirm the conclusion that young children can make more use of their limited information processing capacity in the lesson condition than in the game condition, mainly because there are no secondary goals competing in the lesson condition.

All in all, the conceptual replications of Istomina's study clearly show that young children are better at remembering lists of words in sterile lesson conditions than in more stimulating game conditions, provided that adequate methodological controls are implemented in the experiment. Does this imply that more sterile and—at least to a certain extent—boring learning situations guarantee better memory performance in young children? Is this the message to convey to contemporary memory researchers? We actually do not think so. Let's play the devil's advocate for a moment. We think that the lesson we learned from Istomina's experiment and those replications which corrected its methodological flaws is that the game situation can have enormous motivational effects which in turn seem to facilitate remembering in young children. Apparently, the measures taken to eliminate the experimenter's behavioral differences in the two experimental conditions caused these motivational impacts to disappear. Please note, for example, that in the Hasselhorn and Schneider (Experiment 2) study we did not ask the experimenters to act as enthusiastically in the lesson condition as they already did in the game condition, but instead asked them to slow down and act more neutral in the game condition than they did before. The reason is that we do not know how to elicit the same kind of motivation and a similarly stimulating atmosphere in the lesson condition than is possible in the game condition. We thus cannot exclude the possibility that social context may have a much more tremendous impact on young children's memory than suggested by the replication studies. However, we are left with the dilemma that it is probably impossible to test such an assumption without violating methodological standards typically referred to as internal validity.

In this regard, we would like to note that our inspection of the various replication studies revealed that it is important to distinguish among different groups of context features that may influence remembering. In addition to the environmental features of the context (e.g., game vs. lesson) typically referred to by most researchers discussing the issue of context, it makes sense to include at least two other features, namely features of the learning material (e.g., item attractiveness), and characteristics of the experimenter's behavior as situational features. There is little doubt that these features may influence the outcomes of experimental studies exploring the impact of context in different ways.

In our view, the detailed description of the three replications of Istomina's experiment has revealed that attempts to replicate and validate experimental findings are indeed very important. Given the lack of consistency in findings concerning the effect of context on children's memory, we are afraid that similar problems could arise in replication attempts focusing on other classic developmental studies. Researchers thus are encouraged to try to replicate findings from experimental research. It probably needs several independent replication attempts to get a feeling for the true state-of-the-art concerning

the conceptual issue at stake and to distinguish between obviously solid and more shaky effects. In our case, for example, one of the most solid findings was that the age by activity interaction postulated by Istomina was not obtained in any of the replication experiments. That is, the expectation that recall differences between the game and lesson conditions should be larger for the 4-year-olds than for the 6-year-olds could not be confirmed. Thus the original assumption that a major qualitative change in memory occurs between the ages of 4 and 6 has not been validated by subsequent research.

We think that this is just one illustrative example emphasizing the need for as well as the utility of replication studies in experimental research. We do hope that more researchers and editors of scientific journals will accept the importance of this issue and get more interested in replication studies than they were before. There is no doubt for us that such a revised policy will not cause a delay in scientific progress but instead will not only make possible a faster scientific development but also will add considerably to our knowledge (and not to our prejudices) in various domains.

*We would like to thank Beth Kurtz for her comments on an earlier version of this chapter.

REFERENCES

- Arbinger, R., & Kubsda, J. (1987). Situativer Kontext und Verarbeitungstiefe als Bedingungen für Behaltensleistungen im Vorschulalter. *Psychologie in Erziehung und Unterricht*, *34*, 196–204.
- Baker-Ward, L., Ornstein, P. A., & Holden, D. J. (1984). The expression of memorization in early childhood. *Journal of Experimental Child Psychology*, *37*, 555–575.
- Ceci, S. J., Bronfenbrenner, U., & Baker, J. (1988). Memory development and ecological complexity: The case of prospective remembering. In F. E. Weinert & M. Perlmutter (Eds.), *Memory Development: Universal Changes and Individual Differences* (pp. 243–256). Hillsdale, NJ: Lawrence Erlbaum.
- Cole, M. (Ed.). (1978). *The Selected Writings of A. R. Luria*. New York: Sharpe.
- Finifter, B. M. (1975). Replication and extension of social research through secondary analysis. *Social Science Information*, *14*, 119–153.
- Hasselhorn, M., & Schneider, W. (1990). *External and internal mediators of young children's memory performance*. Unpublished paper. Max Planck Institute for Psychological Research, Munich.
- Istomina, Z. M. (1975). The development of voluntary memory in preschool-age children. *Soviet Psychology*, *13*, 5–64.
- Istomina, Z. M. (1977). The development of voluntary memory in preschool-age children. In M. Cole (Ed.), *Soviet developmental psychology* (pp. 100–159). New York: Sharpe.
- Leontjev, A. N. (1977). *Probleme der Entwicklung des Psychischen*. Kronberg, Germany: Athenäum.
- Meacham, J. A. (1977). Soviet investigations of memory development. In R. V. Kail & J. W. Hagen (Eds.), *Perspectives on the Development of Memory and Cognition* (pp. 273–295). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.
- Neisser, U. (1982). *Memory observed—Remembering in Natural Contexts*. San Francisco: Freeman.

- Perlmutter, M. (1988). Research on memory and its development: Past, present, and future. In F. E. Weinert & M. Perlmutter (Eds.), *Memory Development: Universal Changes and Individual Differences*. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.
- Schneider, W., & Brun, H. (1987). The role of context in young children's memory performance: Istomina revisited. *British Journal of Developmental Psychology*, 5, 333-341.
- Schneider, W., & Pressley, M. (1989). *Memory Development between 2 and 20*. New York: Springer-Verlag.
- Sodian, B., Schneider, W., & Perlmutter, M. (1986). Recall, clustering, and metamemory in young children. *Journal of Experimental Child Psychology*, 41, 395-410.
- Weissberg, J. A., & Paris, S. G. (1986). Young children's remembering in different contexts: A reinterpretation of Istomina's study. *Child Development*, 57, 1123-1129.
- Wertsch, J. V. (1981). *The Concept of Activity in Soviet Psychology*. New York: Sharpe.
- Wippich, W. (1981). Verbessert eine Einkaufssituation die Vorhersage der eigenen Behaltensleistungen im Vorschulalter? *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 8, 280-290.
- Wippich, W. (1985). Untersuchungen zur Entwicklung des Meta-Gedächtnisses in verschiedenen Kontexten. In L. Montada (Ed.), *Bericht über die 7. Tagung Entwicklungspsychologie*. Trier, Germany: Universitätsdruck.