Journal of Educational Psychology
2004, Vol. 96, No. 3, 424–436

0022-0663/04/$12.00 DOI: 10.1037/0022-0663.96.3.424

Activity and Imagined Activity Can Enhance Young Children's Reading Comprehension

Arthur M. Glenberg and Tiana Gutierrez
University of Wisconsin—Madison

Joel R. Levin
University of Arizona

Sandra Japuntich and Michael P. Kaschak
University of Wisconsin—Madison

The Indexical Hypothesis suggests a new method for enhancing children's reading comprehension. Young readers may not consistently "index," or map, words to the objects the words represent. Consequently, these readers fail to derive much meaning from the text. The instructional method involves manipulating toy objects referred to in the text (e.g., a barn, a tractor, a horse, in a text about a farm) to simulate the actions described in the text. Correctly manipulating the objects forces indexing and facilitates the derivation of meaning. Both actual manipulation and imagined manipulation resulted in markedly better (compared with rereading) memory for and comprehension of the text material, thereby lending strong support to the Indexical Hypothesis.

Can young children's reading comprehension be enhanced? Are there potent reading-comprehension strategies that can be identified and prescribed (see, e.g., Ehri, Nunes, Stahl, & Willows, 2001; Symons, McGoldrick, Snyder, & Pressley, 1990)? The Indexical Hypothesis (Glenberg & Kaschak, 2002; Glenberg & Robertson, 2000; Kaschak & Glenberg, 2000) suggests a new sort of answer to these old questions. Because children may not consistently index, or map, written words to the objects the words represent, even when the words are pronounced, these children fail to derive meaning from the text. Consequently, reading becomes an unengaging exercise in word calling. If, as we hypothesize, early young readers' performance can be enhanced by increased indexing, then the following instructional intervention is suggested: While children read texts about events taking place in a particular scenario (e.g., on a farm), objects referred to in the text (e.g., a toy barn, tractor, and horse) are made available, and the children are asked to manipulate those objects to simulate the content of the sentences. Such manipulation should force indexing, thereby facilitating the children's derivation of meaning.

We begin by reviewing the Indexical Hypothesis and some of the research that supports it. The review includes a description of three precedents suggesting that manipulation of objects being read about should enhance children's reading comprehension.[1] We then present three experiments conducted with first- and second-grade readers. These experiments demonstrate large (e.g., 50% and more) positive effects of manipulation on children's recall and application of the material just read. In addition, in Experiment 3, children are trained to imagine manipulating the toys rather than actually manipulating them. This imagined manipulation produces a modest degree of transfer (i.e., strategy maintenance in the absence of instructions). Finally, we contrast the explanation of poor reading comprehension provided by the Indexical Hypothesis with several other accounts based on fluency, inference making, and integration.

Embodied and Nonembodied Theories of Meaning

The Indexical Hypothesis is an embodied (see, e.g., Glenberg, 1997) account of how language becomes meaningful. As such, it is in contrast to most theories of linguistic meaning that are based on amodal, abstract, and arbitrary symbols (AAA symbols). A brief overview of these theories sets the stage for the Indexical Hypoth-

Arthur M. Glenberg, Tiana Gutierrez, Sandra Japuntich, and Michael P. Kaschak, Department of Psychology, University of Wisconsin—Madison; Joel R. Levin, Department of Educational Psychology, University of Arizona.

Michael P. Kaschak is now at the Department of Psychology, Florida State University.

This research was supported by a University of Wisconsin Vilas Associate award to Arthur M. Glenberg, a University of Wisconsin Hilldale award to Tiana Gutierrez, and a National Science Foundation graduate fellowship award to Michael P. Kaschak. Preparation of this article was supported in part by Institute of Education Sciences, Department of Education, Grant R305H030266 to Arthur M. Glenberg and Joel R. Levin. We are very grateful to the students, teachers, and administration at Van Hise and Emerson schools in Madison, Wisconsin. In addition, this article was substantially improved through the peer review process.

Correspondence concerning this article should be addressed to Arthur M. Glenberg, Department of Psychology, University of Wisconsin—Madison, 1202 West Johnson Street, Madison, WI 53706. E-mail: glenberg@wisc.edu

[1] In this article, we generally incorporate Levin's (1986) distinctions in using the term *comprehension* to refer to participants' understanding of text information while a passage is being processed and *memory* and *application* to refer, respectively, to participants' text-based recall/recognition and use/inference subsequent to passage presentation.

esis. One example of a theory making use of AAA symbols is the semantic network (see, e.g., Collins & Loftus, 1975). Consider the node representing the concept "chair." This node is abstract in that it is meant to be the meaning of all chairs, including kitchen chairs, armchairs, and beanbag chairs. The node is amodal in that the same node is contacted whether the chair is seen, talked about, or sat in. Finally, the node is arbitrary in that there is no reason why one node rather than another is used to represent the concept "chair." In fact, because the nodes are arbitrarily related to the concept, meaning is given not by the node itself but by its relations to other nodes such as the furniture node and the object node. Although differing in details, the vast majority of current cognitive theories of meaning use AAA symbols. Thus, theories using propositional representations (see, e.g., Kintsch, 1988) use AAA symbols in that each element in a proposition is equivalent to an AAA symbol. Similarly, most schema theories (e.g., Schank & Abelson, 1977), connectionist theories (e.g., Masson, 1995), and theories using high dimensional spaces (e.g., Burgess & Lund, 1997; Landauer & Dumais, 1997) make use of AAA symbols.

There are several reasons why AAA symbols are predominant. First, these symbols are easy to implement in formal computer and mathematical models. Second, they are absolutely necessary to theories that subscribe to a basic tenet of modern cognitive psychology and artificial intelligence, namely, that thinking is nothing more than the manipulation of AAA symbols by rules (see, e.g., Newell, 1980). If that is the case, then thinking can be adequately modeled (in fact, produced) by computer simulations. In contrast, if thinking requires perceptual information, then theories based on AAA symbols that intentionally strip away all perceptual information and computers that can describe perceptual information but do not literally have any perceptual apparati cannot adequately model human thinking.

Whereas AAA symbols are undeniably popular, there are also several reasons why they are undeniably inadequate. The foremost is the symbol grounding problem (Glenberg & Robertson, 2000; Harnad, 1990; Searle, 1980), namely, that AAA symbols cannot by themselves generate meaning; instead, the symbols must be grounded. Harnad's (1990) version of Searle's (1980) Chinese room argument makes the case cogently. Harnad has one imagine that he or she is a traveler who has landed at a foreign airport equipped solely with a dictionary written in the foreign language the traveler does not speak. Upon encountering a sign at the airport written in the language, the traveler tries to understand the first word by looking it up in the dictionary. Unfortunately, the definition is solely in terms of words in the foreign language that are, to the traveler, prototypical AAA symbols. Undeterred, the traveler looks up the first word in the definition to find that it is defined in terms of other AAA symbols. Of course, no matter how many words the traveler looks up in the dictionary, he or she will never understand any of them, let alone the meaning of the sign at the airport. Yet that is exactly what cognitive theories using AAA symbols expect of people. For example, the nodes in a semantic network are defined solely in terms of their relations to other nodes, just as the words in the dictionary are often defined solely in terms of the other words in the dictionary.

One might suppose that theories making use of AAA symbols have a simple recourse: The AAA symbols are grounded by associating them with perceptual information. Unfortunately, there are four problems with this approach. First, none of the theories actually proposes how this could be done, nor do any of the theories actually make use of perceptual information in accounting for meaning. Second, if perceptual information were to play a major role in producing meaning, then the power of AAA symbols would be severely eroded in that thinking would no longer be due to the manipulation of AAA symbols by rules that can be implemented regardless of the perceptual apparatus (e.g., on a computer). Third, Putnam (1981, and as reviewed in Lakoff, 1987) has proven that when starting with a system of AAA symbols, it is impossible to find the one correct grounding. That is, sets of AAA symbols are the equivalent of a series of algebraic equations in that the numerals, Xs, and Ys in the equation are exactly AAA symbols. A system of equations can be applied to an infinite variety of events in the world. For example, an equation relating speed and time to distance traveled by a car applies to that particular car or any other car, and in fact, it applies to an infinity of moving objects that have the same relations among speed, time, and distance. Similarly, because a set of AAA symbols (e.g., a set of propositions) is equivalent to a set of equations, it applies to an infinite variety of situations—that is, all those situations that have the same set of relations. Consequently, if a person thought in terms of AAA symbols, it would be impossible for that person to know exactly what he or she was thinking about because those same symbols could apply to an infinite number of situations. Fourth, modern neuroscience (e.g., Edelman, 1992) has failed to find anything like AAA symbols in the brain. Rather, perceptual afference and motor feedback seem to be intimately involved in all levels of coding (cf. Pulvermüller, 1999).

The Indexical Hypothesis, like other embodied theories of cognition (Barsalou, 1999; Lakoff, 1987), proposes that meaning requires going beyond AAA symbols. In brief, the hypothesis asserts that the AAA symbols of language (i.e., words), become meaningful by simulating the content of sentences. This simulation is accomplished through three processes. First, words and phrases are indexed to objects in the environment or to perceptual symbols; second, affordances are derived from the objects; third, the affordances are combined, or meshed, as directed by syntax, to produce a coherent simulation.

Consider, for example, how people understand a sentence such as "Art stood on the chair to change the bulb in the ceiling fixture." The first process is that words and phrases, such as *Art* and *the chair* are indexed. Thus, if the sentence were uttered while gesturing to a particular chair, the comprehender would index or map the phrase to that chair (see, e.g., Chambers, Tanenhaus, Eberhard, & Filip, 2002). If the sentence were spoken outside of the context including the chair or if the sentence were read, the comprehender would index the word to a perceptual symbol. Unlike AAA symbols, perceptual symbols (Barsalou, 1999) are not amodal and arbitrary. Instead, they are patterns of neural activity abstracted from perceptual experience by selective attention. That is, these symbols are grounded by their intimate relation with perceptual experience. Thus, as a person attempts to understand the sentence about Art, *Art* may be indexed to a perceptual symbol of a particular person.

The second process is deriving affordances from the indexed perceptual symbols. Affordances (see, e.g., Gibson, 1979) are possibilities for interaction determined by the body and the objects in the environment. Thus, a chair affords sitting for an adult human but not for an infant or a worm. Some chairs afford hiding under for a small child or a worm but not for an adult. That is, affordances are determined both by the object and the biology of the organism interacting with the object.

Finally, affordances are meshed, or integrated, as directed by the syntax of the sentence. We adopt the term *mesh* (in contrast to *combined, associated,* or *integrated*) to highlight a particular characteristic of the presumed process. When affordances are meshed, the combinatory process respects constraints on bodily action. For example, one can mesh the affordances of a chair (it can be stood on) and a light bulb (it can be held) to accomplish the goal of changing the bulb in a ceiling fixture because adult humans can literally combine those actions. However, not all affordances can be combined to accomplish goals. For example, most people would reject as nonsense a sentence such as "Art stood on the can opener to change the bulb in the ceiling fixture." According to the Indexical Hypothesis, this sentence makes little sense because readers have a difficult time imagining how to combine the affordances of a can opener and a light bulb to accomplish the goal of changing a ceiling fixture. Familiarity with prototypical actions (e.g., people often stand on chairs but seldom stand on can openers) is not necessary for mesh. Thus, people find sentences perfectly sensible that describe, for example, Art attempting to change the bulb by standing on an upside-down flowerpot or a TV stand or even by pulling out the drawers of a nearby cabinet, stacking them on the floor, and standing on the stack. That is, when affordances can be meshed, people easily understand language about quite novel events (Glenberg & Robertson, 2000). Kaschak and Glenberg (2000) and Glenberg and Kaschak (2002) have provided a detailed account of how syntactic information can be used to guide the mesh of affordances during sentence understanding.[2]

Three lines of research have provided empirical support for the processes specified by the Indexical Hypothesis. The first line of research has investigated perceptual symbols in conceptual tasks such as feature listing. Barsalou, Solomon, and Wu (1999) demonstrated that variability in the features listed for a concept and variability in the order in which features are listed are controlled by the perceptual simulation being engaged rather than by an abstract, unchanging semantic representation. Using a property verification task, Pecher, Zeelenberg, and Barsalou (2003) observed a priming effect having a perceptual basis. Participants responded whether or not an object (e.g., a lawn mower) has a particular property (e.g., loud). Pecher et al. found that if the perceptual dimension probed on the previous trial (e.g., leaves—rustle) was the same as that probed on the target trial, then responding was faster than if the perceptual dimension probed on the previous trial was different (e.g., leaves—green).

A second line of research has investigated sentence understanding. For example, Zwaan and his associates (e.g., Stanfield & Zwaan, 2001) asked participants to verify that a picture (e.g., of a pencil) depicted an object mentioned in a sentence (e.g., "The pencil is in a cup"). They found that pictures that matched the orientation of the object implied by the sentence (a pencil depicted vertically in this case) were responded to faster than pictures of the object in an orientation that mismatched orientation implied by the sentence (a pencil depicted horizontally). As another example, Glenberg and Kaschak (2002) asked each participant to judge the sensibility of sentences such as "You gave Andy the pizza" or "Andy gave you the pizza" by moving the hand from a start button to a *yes* button. Location of the *yes* button determined if the action of responding "yes" was consistent with the direction implied by the sentence (away from the body for "You gave Andy the pizza") or inconsistent (toward the body). Responding was faster when the hand movement was consistent with the action implied by the sentence than when it was inconsistent. Both sets of results strongly imply that language understanding requires consideration of the perceptual and action characteristics of the situation described.

The third line of research has demonstrated facilitatory effects of action paired with language. It is this line of research that anticipates the effects we report in this article. If language understanding requires indexing of words to objects in the environment or to perceptual symbols, then forcing listeners and readers to complete that process should facilitate comprehension, memory, and application. Consistent with this prediction is a now-standard finding in the memory literature: Memory for a list of tasks (e.g., scratch your ear, break the toothpick, etc.) is greatly enhanced if each task is actually performed in addition to reading the description of the task (Koriat & Pearlman-Avnion, 2003; Nilsson et al., 2000). A finding of Noice and Noice (2001) demonstrates a related effect with discourse. Novice actors were more successful at memorizing dialogue when scripted actions were included than when the actors were simply told to read and memorize the dialogue. Analogous work with young children has supported the beneficial consequences of activity and imagined activity on associative-learning performance (see, e.g., Bender & Levin, 1976; Varley, Levin, Severson, & Wolff, 1974; Wolff & Levin, 1972). Finally, consistent with the earlier findings of Lesgold, Levin, and their colleagues (e.g., Lesgold, DeGood, & Levin, 1977; Lesgold, Levin, Shrimron, & Guttmann, 1975), Rubman and Waters (2000) demonstrated a positive effect of activity/manipulation with third-grade and six-grade students in a text-learning context. Children read a short text and were queried as to whether the text contained any contradictions. Half of the children read the text twice. The others read the text and then used a storyboard to depict it. The children in that latter condition were more successful at detecting errors. Related research and outcomes, based on children's drawing activity, have recently been reported by van Meter (2001).

With this background, we can sketch an answer to the question motivating our research, Can young readers' performance be enhanced? According to the Indexical Hypothesis, an important component of language comprehension is indexing words and phrases to objects or their perceptual symbols. We propose that when young children are first learning oral language, it is in a context that is highly indexed. That is, parents frequently talk about and point to objects in the current environment (e.g., a bottle, a ball, or the baby) and explicitly model actions using gesture, such as waving when requesting that the child "wave bye-bye." Furthermore, parents almost always name and point to novel objects that are introduced to children or that the children are attending to (Masur, 1997). Given the symbol grounding problem discussed earlier, it is hard to imagine how language could be learned

[2] We have discussed the Indexical Hypothesis here only in regard to concrete language and concrete situations. The hypothesis can be extended to language about abstract situations. For discussion, see Glenberg and Kaschak (2002, 2003).

without this sort of indexing. In contrast, when children are learning to read, their attention is focused elsewhere. That is, the child must be closely looking at the written word and directing his or her attention to retrieving the pronunciation of the word or engaging in orthographic-to-phonemic conversion. There is little explicit learning of the connection between the symbol (the written word) and its referent.

A common, albeit implicit, assumption is that young readers do not need to practice indexing written words to their referents. That is, it is assumed that pronouncing the written word aloud is sufficient because it is the same AAA symbol whether heard, produced by the child, or read and spoken aloud. However, to the extent that perceptual symbols must be retrieved from memory, this argument is stretched beyond credibility. There is a vast literature demonstrating strong effects of context on memory retrieval (see, e.g., Morris, Bransford, & Franks, 1977; Tulving & Thomson, 1973), and the context of reading aloud is quite different from the context of hearing speech. For example, speech is often explicitly indexical, accompanied by gesture, and produced and perceived with characteristic rate and intonation. All of these retrieval cues are missing when the child pronounces a word from text. Thus, whereas indexing is a highly practiced skill in the domain of oral language, it may be problematic for young readers. To the extent that the words are not being indexed, reading becomes a meaningless exercise in word calling.

Overview of the Experiments

We used a manipulation procedure to ensure indexing of written words. First- and second-grade children read short texts describing characters and actions in three toy scenarios, a farm, a house, and a gas station/garage. Models of the objects and characters (e.g., for the farm scenario, a barn, a tractor, assorted animals, etc.) were on display in front of the child. After reading each of five selected sentences, the child manipulated the objects to correspond to the sentence. To do this, the child had to index the words and phrases to objects and use the syntax of the sentence to guide manipulation. After reading the text, the toy scenario was covered, and the child was engaged in distracting conversation for 2 min. This distraction was followed by a series of memory and application tests.

In Experiment 1, children were randomly assigned to three groups in which different types of instruction/practice were provided over several sessions: manipulation, read (where children read the texts and observed the scenarios but did not manipulate the toys), and no-practice control. For children in the manipulation group, a fading procedure was used to gradually withdraw the amount of support given to indexing before a transfer test was administered in which children read from a new scenario without benefit of manipulation. Experiment 2 differed from Experiment 1 in several ways. First, there was more practice with manipulation before the strategy-maintenance (transfer) test was administered. Second, the children in a reread group reread each of the critical sentences to help control for the amount of time that children in the manipulate group spent attending to the text. Third, in addition to the memory test of Experiment 1, an application test was included. The application test required the children to draw inferences from what they had read. This test provided another measure of the success of reading-comprehension processes in addition to memory tests. Experiment 3 introduced another change intended to give children practice indexing without explicit manipulation. Children in the imagined manipulation group first practiced physical manipulation, as in the previous experiments. Then, these children were instructed in how to imagine manipulating the objects without physically manipulating them (and the children in the reread group were instructed to reread the critical sentences silently).

Experiment 1

Several questions guided the design of Experiment 1. First, would very young readers be helped by manipulation? Second, would any benefits be moderated by reading ability (see the supplementary analyses following Experiment 3)? Third, would children learn to index as a general strategy such that benefits of manipulation training would transfer to situations in which the children did not physically manipulate? Students participated in Experiment 1 during the fall of their second-grade year. Each child in the manipulate and the read groups participated in six sessions, whereas children in the control group participated in only the first and sixth sessions. The first session was used to obtain a number of measures of students' reading ability. Sessions 2–5 involved training in manipulation and then fading. Session 6 was a test of strategy maintenance.

Method

Participants. Parental permission to participate in the experiment was obtained for 35 children beginning the second grade at a middle-socioeconomic-status public school in Madison, Wisconsin. Three children were eliminated on the basis of extreme difficulty reading a practice text and low scores on the Woodcock (1998) Test of Word Identification. Their raw scores were 5, 13, and 18, whereas the remaining children had scores ranging from 20 to 58 with a mean of 34.5 and a standard deviation of 9.6. In addition, because of absences, data were unavailable for 3 children for Session 6, the maintenance session.

Children were assigned to groups using the following scheme. First, children were grouped by sex. Second, within each sex, children were ranked by performance on the standardized measures of reading given during the first session, and triplets of children with successive ranks were formed. Third, children within a triplet were randomly assigned so that one participated in the manipulate group, one in the read group, and one in the control group. In total, there were 11, 11, and 10 children in the manipulate, read, and control groups, respectively. In all sessions, the interventions and testing were individually administered.

Materials. The scenario toys were commercially available and consisted of a farm scene (including a barn, corral, tractor, several animals, hay, etc.), a house scene (including a house with several rooms and props, a mother, a father, a baby, etc.), and a garage scene (including a garage with elevator, ramps, store, gas pumps, car wash, tow truck, etc.). For each scenario, we wrote five short texts of seven to nine sentences each. A text for the farm scene is given in the Appendix. For each text, we selected five sentences that described actions, and these sentences were followed by "green lights." The green lights were hand-drawn representations of traffic lights with a green light. They were the signal to the child to manipulate the toys in the scenario.

Procedure. The procedure is described for the manipulate group first. The procedure is also summarized in Table 1. Data were collected from each child in six sessions. There were approximately two sessions a week. All sessions were videotaped for later scoring.

Session 1: The child was tested on the Gathercole and Baddeley (1996) Nonword Repetition Test, the Woodcock (1998) Test of Word Identification, and the Woodcock Test of Word Attack Skills (nonword reading). A

Table 1
Procedures for Manipulate and Read Conditions in Experiment 1 During Sessions 2–6

Procedure	Session 2: Training	Session 3: Training	Session 4: Fading	Session 5: Fading	Session 6: Maintenance
Introduction	First scenario	Second scenario			Third scenario
Scenario visible?	Yes	Yes	Yes	No	Yes
First story	Story 1-1[a] (PM/R)[b]	Story 2-1 (PM/R)	Story 1-3 (R/R)	Story 1-4 (R/R)	Story 3-1 (R/R)
	free recall	free recall	free recall	free recall	free recall
	cued recall	cued recall	cued recall	cued recall	cued recall
Second story	Story 1-2 (PM/R)	Story 2-2 (PM/R)	Story 2-3 (R/R)	Story 2-4 (R/R)	Story 3-2 (R/R)
	free recall	free recall	free recall	free recall	free recall
	cued recall	cued recall	cued recall	cued recall	cued recall

[a] The first number indicates the scenario, and the second number indicates the story within the scenario. The actual scenario and story were counterbalanced so that each scenario–story pair appeared approximately equally often in each session and in each position within a session. [b] The first abbreviation indicates the procedure followed in the manipulate condition, and the second abbreviation indicates the procedure followed in the read condition. PM = physical manipulation; R = read.

test of indexing was also given, but because of ceiling effects, these data are not reported.

Session 2: The child was introduced to one of three toy scenes (farm, gas station, or house). Choice of scenario was counterbalanced as best as possible given the sample size. In the introduction, the tester named each object for the child, and each child was required to manipulate the object (e.g., "This is the horse. Put the horse into the corral"). In addition, the child read (with the experimenter's help) a practice text with green-light sentences. Following a green-light sentence, the child was instructed to manipulate the objects in the scene to correspond to the sentence. The child then read two stories referring to the scenario. Choice of stories and their order were counterbalanced. Within each story, five sentences were followed by green lights. Following each story, the toy scene was covered so that it was not visible, and the child was distracted for 2 min with conversation. Then, the child was asked to recall all of the information in the story. Only general prompts were given, such as "What happened next?" Finally, the child was given a cued-recall test for each of the five green-light sentences. The cue was the first part of the sentence, and the child was asked to recall the rest of the sentence.

Session 3: This session was identical to Session 2, except that a different scenario was used.

Session 4: The child read one (new) story from each scenario. There were no green-light sentences (and so no overt manipulation) although the toy scenes were visible while the child read. The point of this session was to determine whether the child would continue to index (as indicated by both looking at the toy scenes and enhanced recall scores compared with the control group) when no manipulation was required.

Session 5: The child read one (new) story from each scenario. The toy scenes were not visible during reading, and there were no green-light sentences. The point of this session was to determine whether the child would index to perceptual symbols (as indicated by recall scores) in the absence of the visible toy scenes.

Session 6: The child was introduced to the third scenario by naming each object and having the child manipulate the object. Then, the child read two stories from the third scenario with no manipulation. Following each story, the free- and cued-recall measures were administered.

Children in the read group were treated exactly like the children in the manipulate group except that there were no green-light sentences during Sessions 2 and 3 (see Table 1). Children in the control group participated in Sessions 1 and 6 during the same times of the semester as did children in the other groups.

Results and Discussion

All statistical analyses were conducted with a Type I error probability set at .05. The data of greatest interest are presented in Table 2. For both free and cued recall, we judged whether or not the main idea of each of the critical sentences was recalled. Because of the similarity of procedures used in Sessions 2 and 3

Table 2
Means (and Standard Deviations) From Experiment 1

Group	Session 2: Training *M* (*SD*)	Session 3: Training *M* (*SD*)	Session 4: Fading *M* (*SD*)	Session 5: Fading *M* (*SD*)	Session 6: Maintenance *M* (*SD*)
Proportion of critical sentences free recalled					
Manipulate	.76 (.20)	.75 (.15)	.55 (.15)	.46 (.10)	.54 (.16)
Read	.49 (.19)	.51 (.20)	.58 (.17)	.64 (.18)	.61 (.14)
Control					.40 (.29)
Proportion of critical sentences cued recalled					
Manipulate	.94 (.07)	.87 (.08)	.79 (.16)	.66 (.14)	.80 (.11)
Read	.77 (.13)	.80 (.18)	.75 (.19)	.79 (.05)	.77 (.18)
Control					.65 (.18)

(see Table 1), the results from these sessions were combined for statistical analysis to increase score stability and associated statistical power. Occasionally, in Experiments 1 and 2, video-recorded responses for a particular story were uninterpretable because participants spoke too softly or because of problems with the video equipment. In these few instances, proportions were estimated by averaging over all relevant stories. For example, if data were missing for one Session 2 story, then the combined Session 2 and 3 score would be based on the three other stories, equally weighted. Estimates were required for two participants (one in the manipulate group and one in the read group) for Session 2.

The proportion of main ideas free recalled from the critical sentences was much greater for the manipulate group (.76) than for the read group (.51), $F(1, 20) = 17.76$, $MSE = .02$, $p < .001$, Cohen's $d = 1.80$. There was a similar effect for cued recall, in that performance in the manipulate group (.90) exceeded that in the read group (.78), $F(1, 20) = 11.68$, $MSE = .01$, $p = .003$, $d = 1.46$. Clearly, manipulation greatly enhanced the children's memory for the text.

Fading began in Session 4. Each child read one text from each of the scenarios introduced in Sessions 2 and 3, but there were no green-light sentences to signal explicit manipulation. As may be seen from Table 2, the difference between the groups vanished, $F < 1$ for both free and cued recall. Apparently, children in the manipulate group had not learned anything that was readily maintained.

In Session 5, the toy scenarios were not visible while the children read. Surprisingly (given the outcomes from the previous sessions), children in the read group outperformed their manipulate counterparts on both performance measures: free recall, $F(1, 20) = 7.62$, $MSE = .02$, $p = .01$, $d = -1.18$; cued recall, $F(1, 20) = 7.32$, $MSE = .01$, $p = .01$, $d = -1.15$. Given other conflicting data from this experiment and the following two experiments, there is no reasonable explanation for these findings.

In Session 6, the children were introduced to a new scenario, and there were no green-light sentences. No obvious benefits of prior manipulation practice were apparent. In particular, including the no-practice control group in an analysis with the two practice groups (manipulate and read) yielded no statistical differences among groups: free recall, $F(2, 26) = 2.59$, $MSE = .04$, $p = .09$; cued recall, $F(2, 26) = 2.41$, $MSE = .03$, $p = .11$.

Given the failure to find statistical differences between the manipulate and the control groups in Session 6, it became important to demonstrate that manipulation, when actually performed, did enhance performance relative to a no-treatment control. To do that, the control group data from Session 6 were compared with the data from Sessions 2 and 3 (combined) for the manipulate and the read groups. Note that this comparison favors the control group because these children (at Session 6) had received several weeks more classroom instruction and practice on reading than the other children (at Sessions 2 and 3). Differences among the three groups emerged for both free recall, $F(2, 27) = 6.66$, $MSE = .05$, $p = .004$, and cued recall, $F(2, 28) = 11.45$, $MSE = .02$, $p < .001$. Fisher least significant difference comparisons revealed that children in the manipulate group remembered more sentences, for both free and cued recall (.76 and .94, respectively), than read participants (.49 and .77, respective ds = 1.19 and 1.26) or children in the control group (.40 and .65, respective ds = 1.56 and 2.13), with no mean differences between read and control groups. Thus, we can be confident that the physical manipulation produced gains relative to both read and control groups.

The results from Experiment 1 are clear. As one would expect from other work showing benefits of action associated with verbal information (see, e.g., Koriat & Pearlman-Avnion, 2003; Nilsson et al., 2000; Noice & Noice, 2001; Rubman & Waters, 2000; van Meter, 2001; Wolff & Levin, 1972), manipulating text referents after reading a sentence greatly enhances memory, even for very young readers. On the other hand, there was no hint of any internalized maintenance of the manipulation strategy. That is, when children in the manipulate group were no longer permitted explicit manipulation, they showed no statistically significant advantage compared with children in either the read (starting in Session 4) or control (in Session 6) groups. Apparently, the children in the manipulate group did not learn either (a) a general skill of indexing or (b) how to apply that skill on their own.

Experiment 2

Experiment 2 was designed to replicate and extend the major results of Experiment 1. It was also designed to test a proposal as to why there was so little Session 6 strategy maintenance in Experiment 1: Two days of manipulation practice (Sessions 2 and 3) may have been insufficient to teach indexing as a skill that children could apply on their own. Thus, in Experiment 2, children in the manipulate group manipulated the toy scenarios for twice as long (Sessions 2–5) before the maintenance test in Session 6. In addition, there were three other important changes from Experiment 1. First, the participants were children starting the second semester of their first grade in school (instead of second graders, as in Experiment 1).[3] Second, beginning in Session 4, the cued-recall test was dropped in favor of a spatial inference test designed to measure application of knowledge gleaned from the text. Whereas the spatial inference test may not be as typical a measure of comprehension or memory as other measures (e.g., retelling of the story), it has the advantage of measuring the extent to which the children successfully combined information explicitly mentioned in the text with information derived from the scenario itself. Furthermore, a justification component on the spatial inference test allowed the children to demonstrate creative application of their comprehension (as described below). Third, children in the reread group were instructed to reread green-light sentences to better control for the amount of time children in the manipulate group spent attending to each of the critical sentences.

Method

Participants. Parental permission to participate in the experiment was obtained for 29 children beginning the second semester of the first grade at the same school as Experiment 1. (One child was eliminated for throwing

[3] We attempted to apply the manipulation intervention to very early readers—that is, children who were just becoming successful as decoders. Thus, in the fall of the school year when we conducted Experiment 1, we chose to work with second-grade children who had generally completed one year of decoding practice the year before; first-grade students would have had very little decoding practice. Experiment 2 was conducted in the second semester. We chose to work with first-grade students who had generally completed one semester of decoding practice.

toys at the experimenter!) Children were assigned to groups using the same scheme as for Experiment 1. In total, there were 8, 9, and 11 children in the manipulate, reread, and control groups, respectively.

Materials. The materials were identical to those used for Experiment 1 except for the spatial inference test questions used in Sessions 4–6. For each text, we wrote three yes/no questions that were difficult or impossible to answer on the basis of the text alone and on the basis of the toy scenario alone. Instead, the correct answer could be derived by integrating the verbal and scenario information. Examples are given in the Appendix. Note that for the first question, the text does not specify Ben's (the farmer's) location. However, there was a hole in the toy hayloft directly above the goat's pen through which a toy bale of hay could be dropped to the goat. Thus the correct answer for this question is "No." The spatial inference questions were not written to probe exclusively knowledge derived from green-light (manipulated) sentences. Hence, they provide a more general measure of text understanding than free or cued recall of the green-light sentences.

After answering each spatial inference question, the child was asked to justify his or her answer. The justification measure was included for two reasons. First, the proportion correct determined by yes/no scoring can be greatly affected by guessing (i.e., the chance probability here is .5). Second, we wanted to give children the opportunity to justify answers based on creative reasoning. For example, a child might say that Ben first walked to the barn and climbed into the hayloft before pushing the hay down the hole. Thus, in the child's mind, Ben and the goat were on the ground floor at the beginning of the story, and the answer to the question should be "Yes." In cases such as this, even though the yes/no answer received a score of 0, the child was awarded credit for the justification, as explained next.

The participants' justifications were evaluated by three people working together. One person was the cuer. This person cued the videotape to the location at which the child was beginning to answer the spatial inference question. The cuing was done out of sight and hearing of the other two people, the scorers. Thus, the scorers remained unaware of the child's group membership. Each scorer evaluated the justifications independently of the other scorer using the following scheme. The justification was given a score of A if the yes/no answer was correct and the child referred to the particulars of the text in justifying the yes/no answer. A score of B was used if the child answered the yes/no question incorrectly but was able to provide a justification that referred to the particulars of the text or a creative extension of the text (e.g., that Ben had been on the first floor of the barn before climbing to the hayloft). A score of C was used if the child provided no justification, if the justification did not refer to the particulars of the text (e.g., "Farmers are never on the same floor as goats"), or if the scorers could not understand the justification. Note that a score of C means unjustified, regardless of the correctness of the yes/no answer. Any disagreements between the two scorers (which occurred on less than 10% of the justifications) were settled by consensus after the scorers recorded their initial evaluations. Less than 5% of the justifications were classified as B. Hence, analyses were conducted on a combined justification score, the sum of the proportions of A and B scores. That is, this combined score is the proportion of questions for which the child provided a reasonable justification for his or her answers.

Procedure. Except in one respect, the procedure for Sessions 1, 2, and 3 were identical to those of Experiment 1 (see Table 3 for a sketch of the procedures). The one difference was that children in the reread group saw the same green lights following sentences as did children in the manipulate group. For children in the reread group, the green light signaled that a sentence was to be reread aloud.

Sessions 4 and 5 were identical to Sessions 2 and 3 except that the spatial inference questions replaced the cued-recall test. Session 6, the strategy maintenance session, was identical to Session 6 in Experiment 1 except, once again, the spatial inference test was administered instead of the cued-recall test.

Results and Discussion

The data of major interest are presented in Table 4. Five recall scores were missing and were estimated from the child's remaining data (three of the estimates were for children in the reread group and two for children in the manipulate group). Three of the spatial inference test justification scores were missing and were estimated from the child's other data. One missing score was in the reread group, and two were in the manipulate group.

The data from Sessions 2 and 3 were again analyzed together to increase score stability and the associated statistical power of the analysis. The mean proportion of green-light sentences freely recalled by children in the manipulate group (.63) exceeded that of children in the reread group (.38), $F(1, 15) = 10.58$, $MSE = .03$, $p = .005$, $d = 1.58$. There was a similar effect for cued recall, $F(1, 15) = 8.52$, $MSE = .02$, $p = .01$, $d = 1.42$, in which the children in the manipulate group (.89) outperformed those in the reread group (.67). Thus, the beneficial effects of manipulation can be found for first graders (this experiment) as well as second graders (Experiment 1).

The data from Sessions 4 and 5 were also analyzed together. For the free-recall test, children in the manipulate group recalled more green-light sentences (.66) than did children in the reread group

Table 3
Procedures for Manipulate and Reread Conditions in Experiment 2 During Sessions 2–6

Procedure	Session 2: Training	Session 3: Training	Session 4: Training	Session 5: Training	Sesion 6: Maintenance
Introduction	First scenario	Second scenario			Third scenario
Scenario visible?	Yes	Yes	Yes	Yes	Yes
First story	Story 1-1[a] (PM/RR)[b]	Story 2-1 (PM/RR)	Story 1-3 (PM/RR)	Story 1-4 (PM/RR)	Story 3-1 (R/R)
	free recall	free recall	free recall	free recall	free recall
	cued recall	cued recall	spatial inference test	spatial inference test	spatial inference test
Second story	Story 1-2 (PM/RR)	Story 2-2 (PM/RR)	Story 2-3 (PM/RR)	Story 2-4 (PM/RR)	Story 3-2 (R/R)
	free recall	free recall	free recall	free recall	free recall
	cued recall	cued recall	spatial inference test	spatial inference test	spatial inference test

[a] The first number indicates the scenario, and the second number indicates the story within the scenario. The actual scenario and story were counterbalanced so that each scenario–story pair appeared approximately equally often in each session and in each position within a session. [b] The first abbreviation indicates the procedure followed in the manipulate condition, and the second abbreviation indicates the procedure followed in the reread condition. PM = physical manipulation; R = read; RR = read and reread critical sentences.

Table 4
Means (and Standard Deviations) From Experiment 2

Group	Session 2: Training *M* (*SD*)	Session 3: Training *M* (*SD*)	Session 4: Training *M* (*SD*)	Session 5: Training *M* (*SD*)	Session 6: Maintenance *M* (*SD*)
			Proportion of critical sentences free recalled		
Manipulate	.55 (.33)	.69 (.34)	.65 (.22)	.68 (.16)	.29 (.26)
Reread	.33 (.24)	.43 (.16)	.32 (.15)	.39 (.20)	.39 (.28)
Control					.44 (.18)
			Proportion of critical sentences correct on spatial inference questions		
Manipulate			.84 (.12)	.85 (.27)	.79 (.21)
Reread			.80 (.14)	.57 (.24)	.65 (.23)
Control					.79 (.21)
			Justification of spatial inference questions		
Manipulate			.78 (.26)	.79 (.29)	.73 (.22)
Reread			.52 (.23)	.48 (.30)	.66 (.22)
Control					.67 (.22)

(.36), $F(1, 15) = 15.46$, $MSE = .03$, $p = .001$, $d = 1.91$. Application, as measured by proportion correct on the dichotomously (yes/no) scored spatial inference test, was somewhat better for the children in the manipulate group (.84) than for the children in the reread group (.69), although this effect was statistically significant only on the basis of a directional test, $F(1, 15) = 3.66$, $MSE = .03$, $p = .04$ (one-tailed), $d = 0.93$. In addition, the difference was statistically significant for the spatial inference justification score, $F(1, 15) = 6.05$, $MSE = .06$, $p = .03$, $d = 1.20$, with children in the manipulation group averaging .78 and those in the reread group averaging .49. These data indicate that manipulation enhanced inference-demanding application in addition to simple memory.

Session 6 was the strategy maintenance session in that a new scenario was introduced and displayed in front of all children, but there were no green-light sentences and no instructions to manipulate the toys. As in Experiment 1, there were no significant differences among the three groups (manipulation, reread, control) for free recall, $F < 1$; proportion correct on the dichotomously scored spatial inference test, $F(2, 25) = 1.17$, $MSE = .05$, $p = .33$; or the justification scores, $F < 1$.

As in Experiment 1, we also included control participants' free-recall data in an analysis of children's Session 2 performance (representing all participants' second test experience with the experimenters and where manipulation was permitted). Although the descriptive statistics again favored manipulate participants (means of .55, .33, and .44 for manipulate, reread, and control, respectively), in contrast to the results of Experiment 1, differences among the three groups were not statistically significant, $F(2, 25) = 1.62$, $MSE = .06$, $p = .22$.[4]

Thus, as in Experiment 1, we have demonstrated that manipulation is very effective when engaged in but that it does not seem to result in a general strategy of indexing that carries over in the absence of direct manipulation, even when manipulation is practiced for four sessions. Nonetheless, the experiment did produce a couple of new pieces of valuable information. First, we uncovered very large effects of manipulation on first-grade readers' free recall, even when manipulation was contrasted with a reread strategy. It is unlikely that the difference was due to time spent on the text. For example, considering the texts of Session 4 (after children had been introduced to both scenarios and had practiced both manipulation and rereading), children spent an average (with standard deviation in parentheses) of 108 (37) s reading and manipulating and 91 (47) s reading and rereading, $t > 1$. Second, we demonstrated that the beneficial effects of manipulation extend to passage-derived application (as measured by spatial inference) as well as memory.

Experiment 3

There are three concerns about the procedures used in Experiments 1 and 2 that may have precluded strategy maintenance in Session 6. First, although manipulation ensured indexing, we did not provide the children with any particular skills or practice that would result in indexing without physical manipulation. Following Wolff and Levin's (1972) approach, however, in Experiment 3, we introduced a new component to the procedure, imagined manipulation, which might provide just such a skill. After practicing physical manipulation, children in the imagined manipulation group were given practice in imagining how they would manipulate the toy scenario. Then, instead of physical manipulation, the green lights signaled that the child should engage in imagined manipulation. In the reread group, the green lights first signaled

[4] On the basis of a reviewer's suggestion to reanalyze the data nonparametrically (presumably because of concerns about small sample sizes and distribution nonnormality), we conducted a set of Kruskal-Wallis rank analysis of variance tests. In those analyses, all statistical decisions were confirmed (along with very similar significance probabilities), with one minor exception: The previously reported statistically significant advantage for manipulate over reread participants on the Sessions 4–5 dichotomous spatial inference measure ($p = .04$, one-tailed) was not strictly reproduced with a corresponding directional test on the mean ranks ($p = .058$, one-tailed).

that the child should reread a sentence aloud. Then, at the point when children in the imagined manipulation group were introduced to imagined manipulation, children in the reread group were taught that the green lights were now a signal to reread sentences silently.

A second reason why the procedures used in Experiments 1 and 2 may have been inadequate is that the benefits of indexing were never explicitly pointed out to the children. That is, they received no explicit metacognitive instruction regarding indexing, whereas work on metacognition as applied to strategy monitoring clearly demonstrates the benefits of this sort of instruction (see, e.g., Duffy, 2002; Ghatala, Levin, Pressley, & Goodwin, 1986; Ghatala, Levin, Pressley, & Lodico, 1985). Consequently, at several points during the procedure, children in the imagined manipulation group were explicitly told that manipulation and imagined manipulation were very beneficial for comprehension and memory, and children in the reread group were told the same for rereading aloud and rereading silently.

A third reason why the procedures used in Experiments 1 and 2 may not have revealed internalized strategy maintenance is that the conditions of the maintenance test may have had too little similarity to the training conditions to have invoked indexing. Consequently, in Session 3 of Experiment 3, we subdivided all participants into two conditions, reminder and no reminder. In the reminder condition, children were introduced to a new scenario, but otherwise, many of the cues and instructions remained the same. That is, green lights were present in the text, and the children were reminded what the green lights signified. In the no-reminder condition, children were introduced to a new scenario, and the green lights were present in the text. However, these children received no instructions or encouragement to apply any strategy.

Method

Participants. Parental permission to participate in the experiment was obtained for 25 children in second semesters of the first and second grades at a low-to-middle-socioeconomic-status public school in Madison. Children from both grades were included to obtain reasonable sample sizes. Several other children were eliminated on the advice of teachers or parents who indicated either that the child was a nonreader or that because of behavior problems, the child would be unlikely to be able to complete a session.

Children were assigned to conditions using a scheme similar to that of Experiments 1 and 2. First, children were grouped by sex. Within each sex grouping, the children were ordered by performance on a text-reading subset of a standardized reading test used in the Madison Metropolitan School District, the Primary Language Arts Assessment (PLAA) test. The PLAA was used instead of the Woodcock test to reduce the amount of time children were taken out of the classroom. Successive groups of three children were randomly assigned so that two were assigned to the imagined manipulation condition and one to the reread condition. Finally, half the children in each group were randomly assigned to the reminder and half to the no-reminder condition. This assignment procedure led to totals of 9, 9, 4, and 3 children in the imagined manipulation/reminder condition, the imagined manipulation/no-reminder condition, the reread/reminder condition, and the reread/no-reminder condition, respectively. We did not expect much of a difference in transfer performance for the two Reread conditions and planned on collapsing the data for statistical analysis, hence the unequal assignment of children to groups. Also, because of limited student resources, we decided not to include a no-practice control condition here.

Materials. The materials for the farm and house scenarios were used in counterbalanced order so that approximately half the children in each group had each scenario for the maintenance test. In addition, several other sets of materials were written to provide practice in imagined manipulation and rereading silently and to encourage the use of these strategies.

Procedure. The procedure for the imagined manipulation condition is described first. Table 5 provides a summary of the procedures. In Session 1, children were introduced to one of the scenarios and the green lights. However, there was more explicit instruction as to the helpfulness of the manipulation strategy. The child read the first story using physical manipulation. Then, the scenario toys were covered, and the child was distracted for 2 min. The free-recall and spatial inference tests followed the distraction. Next, the tester again noted the effectiveness of manipulation and introduced the idea of imagined manipulation as being equally effective. The following are verbatim instructions regarding imagined manipulation for children reading stories for the farm scenario.

Table 5
Procedures for Imagined Manipulate and Reread Conditions in Experiment 3

Procedure	Session 1: Training	Session 2: Training	Session 3: Maintenance
Introduction	First scenario		Second scenario
Scenario visible?	Yes	Yes	Yes
First story	Story 1-1[a] (PM/RR)[b] free recall spatial inference test	Story 1-3 (IM/SR) free recall spatial inference test	Reminder: Story 2-1 (IM/SR) No reminder: Story 2-1 (R/R) free recall spatial inference test
Additional training	Training on IM or SR		
Second story	Story 1-2 (IM/SR) free recall spatial inference test	Story 1-4 (IM/SR) free recall spatial inference test	Reminder: Story 2-2 (IM/SR) No reminder: Story 2-2 (R/R) free recall spatial inference test

[a] The first number indicates the scenario, and the second number indicates the story within the scenario. The actual scenario and story were counterbalanced so that each scenario–story pair appeared approximately equally often in each session and in each position within a session. [b] The abbreviation indicates the procedure followed in the manipulate condition, and the second abbreviation indicates the procedure followed in the read condition. PM = physical manipulation; R = read; RR = read and reread critical sentences; IM = imagined manipulation; SR = silent rereading (following reading aloud).

> Suppose that you read, "The goat chased the horse into the corral." If you were to act out the sentence, you would move the goat out of his pen and run to the horse. Then, the horse would jump into the corral. Now, instead of acting out the sentence IMAGINE how you would move the goat to the horse and IMAGINE how the horse would jump into the corral. Don't really act out the sentence, but IMAGINE how you would move the toys to act out the sentence. Often, it will help if you read the sentence and then LOOK at the toys to help you to IMAGINE how you would act out the sentence.
>
> There is one other thing I would like you to do. When you finish reading the sentence, put your finger on the green light. Putting your finger on the light is a reminder to IMAGINE how you would move the toys to act out the sentence. So, read the sentence out loud, put your finger on the green light, look at the toys, and then IMAGINE acting out the sentence. Imagining will help you to remember the sentences and understand the story.

Children then practiced reading and imagining for two sentences. After each sentence, the tester asked the child to describe the image that she or he had constructed. In particular, the child was prompted to describe intermediate steps between the information mentioned explicitly in the sentence. Finally, the child read a second story using the imagined manipulation strategy and, after distraction, recalled the story and answered the spatial inference questions. The session ended with further discussion of the efficacy of imagined manipulation.

Session 2 consisted of the reading of two more stories using imagined manipulation. Each story was followed by both the free-recall and spatial inference tests.

Session 3 differed for children in the reminder and no-reminder conditions. For the reminder condition, children were introduced to the characters in the new scenario and reminded about using the green lights as a signal to imagine manipulation. Then, the children read and were tested on two stories from this scenario. For the no-reminder condition, children were introduced to the characters of the new scenario. Then, although the green lights were present, no mention was made of them, nor were children reminded about imagined manipulation. These children also read two stories from the new scenario.

For children in the two reread conditions, the sessions were identical to those described above except that these children practiced rereading aloud or silently whenever children in the imagined manipulation condition practiced physical manipulation or imagined manipulation, respectively. Children in the reread conditions were also told about the efficacy of rereading.

Results

The results of major interest are presented in Table 6. There were no missing data requiring estimation for the recall scores or the spatial inference test scores. We consider first performance on the first story in Session 1. The contrast of manipulation and rereading provided a replication of the conditions in the first two experiments. Compared with the children in the reread condition, the children in the manipulation condition recalled a greater proportion of the critical sentences, $F(1, 23) = 9.80$, $MSE = .06$, $p = .005$, $d = 1.39$, and answered correctly a greater proportion of the dichotomous spatial inference questions, $F(1, 23) = 4.32$, $MSE = .03$, $p = .05$, $d = 0.81$, although the effect was not statistically significant on the inference justification measure, $F(1, 23) = 1.93$, $MSE = .07$, $p = .18$.

Performance on the second story in Session 1 and the two stories in Session 2 provided new information: a comparison of imagined manipulation with rereading silently. The data from these three stories were collapsed for analysis, and the analyses support the claim that imagined manipulation results in stronger memory and better application than rereading. Children in the imagined manipulation group recalled a greater proportion of the critical sentences (.66) than did children in the reread group (.27), $F(1, 23) = 17.56$, $MSE = .04$, $p < .001$, $d = 1.87$, and children in the imagined manipulation group answered correctly a greater proportion of the spatial inference questions (.87) than did children in the reread group (.67), $F(1, 23) = 11.29$, $MSE = .02$, $p < .005$, $d = 1.50$. In addition, children in the imagined manipulation group provided somewhat stronger justifications (.72) than did children in the reread group (.52), $F(1, 22) = 3.80$, $MSE = .05$, $p < .05$ (one-tailed), $d = 0.88$. This comparison is statistically significant in an analysis of Session 2 performance alone, $F(1, 22) = 7.03$, $MSE = .05$, $p < .02$, $d = 1.19$. As in Experiment 2, it is unlikely that the difference was due to time spent on the text. For example, con-

Table 6
Means (and Standard Deviations) From Experiment 3

Group	Session 1: Story 1 *M* (*SD*)	Session 1: Story 2 *M* (*SD*)	Session 2 *M* (*SD*)	Session 3: Maintenance *M* (*SD*)
Proportion of critical sentences free recalled				
Imagined manipulation	.62 (.23)[a]	.66 (.24)	.66 (.25)	.72 (.19)
Reread	.29 (.28)	.31 (.23)	.24 (.18)	.46 (.28)
Proportion correct on spatial inference questions				
Imagined manipulation	.93 (.14)[a]	.83 (.21)	.89 (.14)	.86 (.14)
Reread	.76 (.25)	.72 (.23)	.64 (.20)	.76 (.13)
Justification of spatial inference questions				
Imagined manipulation	.67 (.25)[a]	.66 (.31)	.76 (.18)	.81 (.20)
Reread	.50 (.32)	.57 (.46)	.49 (.31)	.61 (.22)

[a] For Story 1, this represents a physical manipulation condition.

sidering the texts of Session 2, children spent an average (with standard deviation in parentheses) of 84 (41) s reading and imagining manipulation and 71 (48) s reading and silently rereading, $t > 1$.[5] Data from Session 3 determine whether or not there was strategy maintenance when a new scenario was introduced without extensive prompting to use any strategy. Because of reviewer concerns about the small sample sizes in relation to detecting an interaction between the reminder and strategy factors, presentation of the Session 3 data and discussion focus on strategy comparisons collapsed across the reminder factor.[6] For free recall of the green-light sentences, there was a main effect of strategy, $F(1, 23) = 7.62$, $p = .01$, $d = 1.23$, with children in the imagined manipulation group recalling more than reread participants (.72 vs. .46, respectively). There was no statistically significant effect on the dichotomous spatial inference justification measure, $F(1, 23) = 2.54$, $p = .13$. However, there was a significant effect of strategy on the spatial inference justification measure, $F(1, 23) = 4.45$, $p = .05$, $d = 0.94$.

Supplementary Analyses

Might any benefits of manipulation reflect an increase in fluency rather than indexing? That is, as the children became more familiar with the toys and their names (because of manipulation), they read the texts more fluently, overcame working memory limitations (see, e.g., Perfetti, 1985), and learned more. There are two pieces of data that speak against a simple fluency account. First, we tracked the number of miscues (i.e., mispronunciations or failures to produce a word) and found that there were no differences among conditions in number of miscues in Sessions 1–3, all Fs < 1. Second, for children in the reread group, we measured whether or not fluency increased on rereading aloud the green-light sentences in the first story in Session 1. Although fluency increased on 89% of the reread sentences, as was noted in the Session 1 analyses, performance by children in the reread group was markedly inferior to that of children who either physically (Story 1) or mentally (Story 2) manipulated the story objects.

Additional analyses were conducted to determine if there were Aptitude × Treatment interactions. That is, did manipulation differentially benefit children who were particularly poor (or good) readers (as reflected by their Woodcock and PLAA scores)? These analyses were conducted for each experiment separately as well as by combining the data across the experiments. No interactions between children's initial reading ability and experimental condition were detected. That is, the regression of children's reading-task performance on their reading ability did not differ statistically by experimental condition (although this conclusion must be tempered by the relatively small sample sizes associated with these analyses).

General Discussion

The present set of experiments extends the findings from earlier associative-learning research to suggest that both manipulation and imagined manipulation can greatly enhance young children's reading performance, as reflected by both their memory for what they have read and their ability to derive text-based inferences. That manipulation-versus-read/reread differences were generally statistically significant here, even with such small sample sizes and (relatively) unselected readers, is particularly noteworthy. From a practical standpoint, these differences are substantial in magnitude. For example, across experiments, the average free-recall facilitation resulting from physical manipulation amounted to 78%, and the average improvement due to imagined manipulation was 99%. Nonetheless, additional research is needed to demonstrate that practice with imagined manipulation leads to long-term maintenance and transfer in situations that are farther removed from the experimental situation. For example, research by Ghatala, Levin, and Pressley (e.g., Ghatala et al., 1985) has strongly indicated that when it comes to learning-strategy maintenance, explicit performance-based feedback following experience with both effective and ineffective strategies is a critical element.

The Indexical Hypothesis provided the theoretical background that guided this research. In brief, that hypothesis asserts that meaning arises from simulating the content of sentences. This simulation requires indexing words to the objects and actions those words represent, deriving affordances (how those objects can be manipulated), and meshing those affordances as directed by the syntax of the sentence. The manipulation condition required children to explicitly index the words to objects and to align the objects as directed by the syntax of the sentence. Thus, the manipulation condition guaranteed meaningful comprehension as described by the Indexical Hypothesis.

Other approaches to language comprehension may also apply to our results. Here, we consider two such approaches, one based on fluency and one based on inference making and integration. We consider the fluency approach first. Among others, Perfetti (1985) and Rayner, Foorman, Perfetti, Pesetsky, and Seidenberg (2001) have emphasized the need to develop fluency in going from an orthographic code (the written words) to a phonological representation. We agree that this is a critical step in reading. In addition, as we discussed earlier, if derivation of the phonological representation is slow or results in a prosodically awkward (e.g., poorly accented) representation, then indexing will be difficult. Nonetheless, fluency by itself is insufficient to account for our data. That is, if the manipulation condition generated better comprehension, memory, and application through enhanced fluency, then we would have expected to see fewer miscues in this condition compared with the reread condition. That was not the case. Furthermore, we documented enhanced fluency in the reread group when children reread sentences aloud for a second time (Experiments 2 and 3). Nonetheless, performance in this condition was markedly inferior to performance in the manipulate condition.

A second approach that is more compatible with our data is that manipulation helps children to derive inferences necessary to construct integrated mental models (Mayer, 1989; Moreno & Mayer, 1999; Rubman & Waters, 2000). A mental model is often conceptualized as a representation that goes beyond information given explicitly in the text by incorporating inferences and world knowledge. Through these processes, the mental model becomes a

[5] The data reported in the text exclude one outlier time in the reread condition. With that child's data included, the mean and standard deviation are 104 (97) s, indicating that the average time in the reread condition was numerically greater than in the imagined manipulate condition.

[6] The main effect conclusions based on initially conducted 2 × 2 factorial analyses are completely consistent with those presented here.

representation of what the text is about rather than a representation of the text itself (see, e.g., Glenberg, Meyer, & Lindem, 1987). The Indexical Hypothesis adds several new ideas to accounts of mental models. First, according to the Indexical Hypothesis, mental models are constructed from meshing affordances derived from perceptual symbols, rather than AAA symbols. Second, indexing words to specific objects or specific perceptual symbols (e.g., this particular barn with a hole in the hayloft rather than barns in general) is an important type of inference that goes beyond the text. That is, once a word is indexed, the affordances of the indexed object are a source of critically important information (e.g., the fact that the hole in the hayloft is large enough for the bale of hay to fit through it and that the hole is situated over the goat's pen so that the hay falls into the pen). Third, the Indexical Hypothesis suggests why building mental models through manipulation is a particularly effective strategy for early readers. Namely, manipulation ensures indexing. Thus, the Indexical Hypothesis is generally consistent with claims of mental model theory (a representation based on extratextual knowledge) but provides a different specification of components of the model, namely, meshed affordances.

In conclusion, we have demonstrated how an embodied approach to language comprehension, the Indexical Hypothesis, can be applied to enhance early reading performance. The hypothesis offers an approach to language comprehension that suggests a powerful faded teaching technique, manipulation and imagined manipulation. Furthermore, the hypothesis provides a new perspective on research demonstrating the importance of inference making and the construction of mental models.

References

Barsalou, L. W. (1999). Perceptual symbols systems. *Behavioral and Brain Sciences, 22,* 577–660.

Barsalou, L. W., Solomon, K. O., & Wu, L.-L. (1999). Perceptual simulation in conceptual tasks. In M. K. Hiraga, C. Sinha, & S. Wilcox (Eds.), *Cultural, typological, and psychological perspectives in cognitive linguistics: The proceedings of the 4th conference of the International Cognitive Linguistics Association* (Vol. 3, pp. 209–228). Amsterdam: John Benjamins.

Bender, B. G., & Levin, J. R. (1976). Motor activity, anticipated motor activity, and young children's associative learning. *Child Development, 47,* 560–562.

Burgess, C., & Lund, K. (1997). Modelling parsing constraints with high-dimensional context space. *Language and Cognitive Processes, 12,* 177–210.

Chambers, C. G., Tanenhaus, M. K., Eberhard, K. M., & Filip, H. (2002). Circumscribing referential domains during real-time language comprehension. *Journal of Memory and Language, 47,* 30–49.

Collins, A. M., & Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychological Review, 82,* 407–428.

Duffy, G. G. (2002). The case for direct explanation of strategies. In C. C. Block & M. Pressley (Eds.), *Comprehension instruction* (pp. 28–41). New York: Guilford Press.

Edelman, G. M. (1992). *Bright air, brilliant fire.* New York: Penguin.

Ehri, L. C., Nunes, S., Stahl, S., & Willows, D. (2001). Systematic phonics instruction helps students learn to read: Evidence from the National Reading Panel's meta-analysis. *Review of Educational Research, 71,* 393–447.

Gathercole, S. E., & Baddeley, A. D. (1996). *The Children's Test of Nonword Repetition*. London: Psychological Corporation.

Ghatala, E. S., Levin, J. R., Pressley, M., & Goodwin, D. (1986). A componential analysis of the effects of derived and supplied strategy utility information on children's strategy selections. *Journal of Experimental Child Psychology, 41,* 76–92.

Ghatala, E. S., Levin, J. R., Pressley, M., & Lodico, M. G. (1985). Training cognitive-strategy monitoring in children. *American Educational Research Journal, 22,* 199–215.

Gibson, J. J. (1979). *The ecological approach to visual perception*. New York: Houghton Mifflin.

Glenberg, A. M. (1997). What memory is for. *Behavioral and Brain Sciences, 20,* 1–55.

Glenberg, A. M., & Kaschak, M. P. (2002). Grounding language in action. *Psychonomic Bulletin and Review, 9,* 558–565.

Glenberg, A. M., & Kaschak, M. P. (2003). The body's contribution to language. In B. Ross (Ed.), *The psychology of learning and motivation* (Vol. 43, pp. 93–126). New York: Elsevier Science.

Glenberg, A. M., Meyer, M., & Lindem, K. (1987). Mental models contribute to foregrounding during text comprehension. *Journal of Memory and Language, 26,* 69–83.

Glenberg, A. M., & Robertson, D. A. (2000). Symbol grounding and meaning: A comparison of high-dimensional and embodied theories of meaning. *Journal of Memory and Language, 43,* 379–401.

Harnad, S. (1990). The symbol grounding problem. *Physica D: Nonlinear Phenomena, 42,* 335–346.

Kaschak, M. P., & Glenberg, A. M. (2000). Constructing meaning: The role of affordances and grammatical constructions in sentence comprehension. *Journal of Memory and Language, 43,* 508–529.

Kintsch, W. (1988). The role of knowledge in discourse comprehension: A construction–integration model. *Psychological Review, 95,* 163–182.

Koriat, A., & Pearlman-Avnion, S. (2003). Memory organization of action events and its relationship to memory performance. *Journal of Experimental Psychology: General, 132,* 435–454.

Lakoff, G. (1987). *Women, fire, and dangerous things: What categories reveal about the mind*. Chicago: University of Chicago Press.

Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review, 104,* 211–240.

Lesgold, A. M., DeGood, H., & Levin, J. R. (1977). Pictures and young children's prose learning: A supplementary report. *Journal of Reading Behavior, 9,* 353–360.

Lesgold, A. M., Levin, J. R., Shimron, J., & Guttmann, J. (1975). Pictures and young children's learning from oral prose. *Journal of Educational Psychology, 67,* 636–642.

Levin, J. R. (1986). Four cognitive principles of learning-strategy instruction. *Educational Psychologist, 21,* 3–17.

Masson, M. E. J. (1995). A distributed memory model of semantic priming. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21,* 3–23.

Masur, E. F. (1997). Maternal labeling of novel and familiar objects: Implications for children's development of lexical constraints. *Journal of Child Language, 24,* 427–439.

Mayer, R. E. (1989). Systematic thinking fostered by illustrations in scientific text. *Journal of Educational Psychology, 81,* 240–246.

Moreno, R., & Mayer, R. E. (1999). Cognitive principles of multimedia learning: The role of modality and contiguity. *Journal of Educational Psychology, 91,* 358–368.

Morris, C. D., Bransford, J. D., & Franks, J. J. (1977). Levels of processing versus transfer appropriate processing. *Journal of Verbal Learning and Verbal Behavior, 16,* 519–533.

Newell, A. (1980). Physical symbol systems. *Cognitive Science, 4,* 135–183.

Nilsson, L.-G., Nyberg, L., Klingberg, T., Aberg, C., Persson, J., & Roland, P. E. (2000). Activity in motor areas while remembering action events. *NeuroReport, 11,* 2199–2201.

Noice, H., & Noice, T. (2001). Learning dialogue with and without movement. *Memory & Cognition, 29,* 820–827.

Pecher, D., Zeelenberg, R., & Barsalou, L. W. (2003). Verifying properties from different modalities for concepts produces switching costs. *Psychological Sciences, 14,* 119–124.

Perfetti, C. A. (1985). *Reading ability*. New York: Oxford University Press.

Pulvermüller, F. (1999). Words in the brain's language. *Behavioral and Brain Sciences, 22,* 253–336.

Putnam, H. (1981). *Reason, truth, and history*. Cambridge, England: Cambridge University Press.

Rayner, K., Foorman, B. R., Perfetti, C. A., Pesetsky, D., & Seidenberg, M. S. (2001). How psychological science informs the reaching of reading. *Psychological Science in the Public Interest, 2,* 31–74.

Rubman, C. N., & Waters, H. S. (2000). A, B, seeing: The role of constructive processes in children's comprehension monitoring. *Journal of Educational Psychology, 92,* 503–514.

Schank, R. C., & Abelson, R. P. (1977). *Scripts, plans, goals, and understanding: An inquiry into human knowledge structures.* Hillsdale, NJ: Erlbaum.

Searle, J. R. (1980). Minds, brains and programs. *Behavioral and Brain Sciences, 3,* 417–424.

Stanfield, R. A., & Zwaan, R. A. (2001). The effect of implied orientation derived from verbal context on picture recognition. *Psychological Science, 12,* 153–156.

Symons, S., McGoldrick, J. A., Snyder, B. L., & Pressley, M. (1990). Reading comprehension. In M. Pressley & associates (Eds.), *Cognitive strategy instruction that really improves children's academic performance* (pp. 45–70). Cambridge, MA: Brookline.

Tulving, E., & Thomson, D. M. (1973). Encoding specificity and retrieval processes in episodic memory. *Psychological Review, 80,* 352–373.

van Meter, P. (2001). Drawing construction as a strategy for learning from text. *Journal of Educational Psychology, 93,* 129–140.

Varley, W. J., Levin, J. R., Severson, R. A., & Wolff, P. (1974). Training imagery production in young children through motor involvement. *Journal of Educational Psychology, 66,* 262–266.

Wolff, P., & Levin, J. R. (1972). The role of overt activity in children's imagery production. *Child Development, 43,* 537–547.

Woodcock, R. W. (1998). *Woodcock Reading Mastery Tests—Revised*. Circle Pines, MN: American Guidance Service.

Appendix

Example Text (From Farm Scenario) and Spatial Inference Questions

Title: Breakfast on the Farm

Ben needs to feed the animals.
He pushes the hay down the hole.*
The goat eats the hay.*
Ben gets eggs from the chicken.*
He puts the eggs in the cart.*
He gives the pumpkins to the pig.*
All the animals are happy now.

Spatial Inference Questions (Experiments 2 and 3)

At the beginning of the story, is Ben on the same floor as the goat?
When Ben is giving the pig the pumpkins, can he see the sheep?
Was the cart next to Ben when he got the eggs?

*Sentence followed by a green light.

Received July 18, 2003
Revision received February 3, 2004
Accepted February 17, 2004 ■