Working Memory, Strategy Knowledge, and Strategy Instruction in Children With Reading Disabilities

H. Lee Swanson,1 Pam Kehler,2 and Olga Jerman3

Abstract
Two experiments investigated the effects of strategy knowledge and strategy training on the working memory (WM) performance in children (ages 10–11) with and without reading disabilities (RD). Experiment 1 examined the relationship between strategy knowledge (stability of strategy choices) and WM performance as a function of initial, gain (cued), and maintenance conditions. WM performance was significantly improved for both groups under cued conditions; however, the performances of children with RD were inferior to those of children without RD across all memory conditions. Measures of WM capacity rather than strategy stability or processing efficiency best predicted reading comprehension performance. Experiment 2 assessed the effects of strategy training on WM performance by randomly assigning children to strategy instruction or control conditions. Significant improvements in WM performance occurred as a function of training conditions, but the residual WM differences between the reading groups remained. Although the results showed that stable strategy choices, cued performance, and strategy instruction significantly bolstered WM performance in children with RD, their overall WM performance, however, was constrained by capacity limitations.

Keywords
reading disabilities, working memory, strategies, strategy knowledge

Working memory (WM), the limited capacity system that allows simultaneous storage and processing of temporary information (Baddeley & Logie, 1999), has been the focus of several studies related to reading disabilities (RD) in children (e.g., Passolunghi, Cornoldi, & De Liberto, 1999; Siegel & Ryan, 1989; Swanson, Ashbaker, & Lee, 1996). Several studies suggest that children with RD suffer WM deficits related to the phonological loop, a component of WM that specializes in the retention of speech-based information (e.g., Siegel & Ryan, 1989; Swanson & Siegel, 2001). Children with RD also have deficits related to controlled attentional processing (e.g., maintaining task-relevant information in the face of distraction or interference), a component of the executive system of WM, that is independent of their problems in the phonological system (Chiappe, Hasher, & Siegel, 2000; De Beni, Palladino, Pazzaglia, & Cornoldi, 1998; De Jong, 1998; Gathercole, Alloway, Willis, & Adams, 2006; Swanson, 1993a, 1993b). One factor that has been overlooked in linking WM to reading performance in children with RD is their knowledge and utilization of strategies. Research is fairly clear that children with RD fail to implement memory strategies spontaneously and efficiently (see O’Shaughnessy & Swanson, 1998, for a meta-analysis of this literature); however, it is unclear whether such children’s knowledge base about strategies is deficient and/or if such children can be taught to use strategies on WM tasks to improve performance. For the present study, we designed two experiments to determine whether knowledge about strategies and/or strategy utilization may underlie some of the relationship between WM and reading performance reported in the literature.

Strategy Knowledge and WM Performance
The general importance of strategy knowledge has been shown in the child development research on metamemory (see Schneider & Pressley, 1997, for a review). Strategy knowledge has been shown to improve with age (Justice, 1985; Kreutzer, Leonard, & Flavell, 1975) and is related to

1University of California, Riverside
2Riverside School District
3Frostig School, Pasadena

Corresponding Author:
H. Lee Swanson, Graduate School of Education, University of California, Riverside, 92521
E-mail: Lee.Swanson@ucr.edu
strategy choices matched the demands of the task. Based on these findings, Swanson argued that children with RD do not suffer deficits in strategy knowledge but rather have constraints in WM capacity.

The purpose of Experiment 1 was to determine the influence of strategy knowledge on the WM performance of children with RD. One hypothesis explored is that children with RD differ in WM deficits, and these deficits can be partly accounted for by the children's knowledge and consistent use of effective strategies. This hypothesis is drawn from Coyle, Read, Gaultney, and Bjorklund (1998), who compared strategic knowledge and memory performance in 9-year-old gifted and nongifted children. They found that strategic variability (fluctuation) across memory trials was related to memory performance. Stability in strategy use, not the particular strategy selected, was associated with high levels of recall for gifted, but not for nongifted, children. Thus, although there were weak links between the type of strategy selected and memory performance, the stability of choice was related to recall performance in the gifted sample. Based on this research, Experiment 1 explores the possibility that the relationship between strategy choice and WM performance is less stable in children with RD than without RD.

To this end, children in Experiment 1 were administered verbal and visual-spatial WM tasks. For each task, participants were asked to select from an array of strategies the strategy they believe will help their retrieval. If strategy knowledge underlies WM performance, then children with RD would be expected to have less stable strategy choices than do children without RD. Assuming that strategy choices may only partly account for the residual differences between reading groups, however, Experiment 1 assessed whether the relationship between WM and reading relates to individual variations in processing efficiency and/or processing capacity.

To assess processing efficiency and/or processing capacity, this study presented WM tasks under authentic conditions that enhanced performance to the participant's maximum span length. Cues helped the participant reinstate the memory trace and/or retrieve forgotten items. The cueing conditions maximized processing efficiency by bringing an individual's WM performance to an asymptotic level. The number of cues to achieve an asymptotic level serves as an indirect measure of processing efficiency (i.e., fewer cues relate to greater efficiency). The WM tasks were also presented at the asymptotic level (the maximum span length established with cues) but without cues. This condition included the same WM tasks that matched each participant's highest WM span level. Thus, each participant was presented items calibrated to his or her asymptotic level of WM performance. This calibrating allowed for the assessment of processing constraint beyond the learning of items. The ability to maintain a high level of performance without cues serves as an indirect index of demands on WM capacity.

**Strategy Training and WM Performance**

Experiment 2 focused on the effects of strategy training on the relationship between WM and reading performance. Investigated in this experiment was whether strategy training reduces the residual variance that exists in WM performance between children with and without RD. The hypothesis under investigation is that children with RD (children with low WM spans) benefit more from strategy instruction than do children without RD (children with high spans) because they are more likely to experience greater processing constraints on WM tasks than are children without RD. In one of the few studies on the effects of strategy training on WM performance in children, Klingberg et al. (2005) found that when children with attention-deficit/hyperactivity disorder (ADHD) were exposed to a computerized WM training program, significant improvements emerged on measures of verbal and visual-spatial memory and complex reasoning (Raven Colored Progressive Matrices Test) relative to the control conditions. Improvements in WM and their links to reasoning were attributed to activities of the central executive system (e.g., response inhibition). However, the study did not address whether the treatment effects moderated the covariation between WM performance and academic achievement (e.g., reading comprehension). The only studies we are aware of that address
this issue involve college students (McNamara & Scott, 2001; Turley-Ames & Whitfield, 2003). Both of these studies found that strategy intervention improved posttest WM performance and moderated the correlation between WM and reading comprehension. For example, Turley-Ames and Whitfield showed that the correlations between reading and WM were higher in the pretest than in the posttest condition, suggesting that strategy instruction reduced the variance related to individual differences at posttest. The correlations were higher at posttest than at pretest for certain strategy training conditions. This finding suggests that controlling for strategy use may have removed error variance (due to variability in strategy use), which in turn provided a better measure of individual differences in WM. Overall, these results were interpreted as suggesting that strategy training helped allocate WM resources more efficiently in low-span participants (see Turley-Ames & Whitfield, 2003, for further discussion of this hypothesis). In line with the findings of Turley-Ames and Whitfield, Experiment 2 in the present study tests the hypothesis that children with RD, because of their low WM span, benefit more from strategy instruction than do children without RD (children with high spans).

A competing hypothesis is that although strategy training may improve performance in children with RD, it plays a minor role in reducing the variance between WM and reading performance. For example, Kane, Conway, Hambrick, and Engle (2007) in their analysis of WM strategy training studies (i.e., McNamara & Scott, 2001; Turley-Ames & Whitfield, 2003), indicated that these studies “do not contribute to the correlations between WM span and higher-order abilities” (p. 31). They found that the standard deviations in WM span were larger after training, rather than before training, suggesting that span variability was increased rather than reduced after training. They concluded that although strategies may improve performance, they do not contribute to the correlations between WM span and achievement. Thus, an alternative hypothesis tested in Experiment 2 is whether the correlations between WM and reading performance remain stable from pretest to posttest. A finding that the correlations remain stable and/or increase would suggest that processing constraints, rather than processing efficiency, underlie the correlations between WM and reading (cf. Kane et al., 2007, p. 31).

**Experiment 1**

Experiment 1 investigates whether the poor WM performance in children with RD relates to strategy knowledge, processing efficiency, and/or constraints in WM capacity. To examine these possibilities, verbal and visual-spatial WM tasks were presented under three conditions: (a) without cues to assess initial performance, (b) with cues to optimize performance (considered as gain or asymptotic performance), and (c) without cues on the highest level of performance achieved under gain (cued) conditions (maintenance).

The logic for the conditions is as follows (also see Swanson, 1999, 2003). The noncued or initial condition reflects the baseline for each participant’s self-initiated processes to access information. The cueing condition enhances the access to stored items by tailoring cues to help participants reinstate memory traces or to retrieve forgotten items from the initial (or baseline) conditions. Previous studies have shown that the cueing conditions improve performance by as much as 1 standard deviation (Swanson, 1992, 1993b). This occurs because the systematic cueing procedures emphasize sequential processing strategies and thereby reduce the number of competing strategies employed. If the locus of WM problems is in the retrieval phase, one would expect a reduction in reading group difference for this condition when compared to the initial condition. One reservation in arguing that performance differences between RD and skilled readers are due to enhanced retrieval efficiency (i.e., improved access to items previously forgotten) is that the manipulations between the noncued (initial) and cueing condition are limited to individual differences in the preservation of information during processing. Thus, the highest level achieved under the cueing condition is readministered (referred to as the maintenance condition) using the same materials but without cues to support performance. Calibrating this condition allows us to capture processing differences between groups beyond the learning of items. Because each participant is presented items calibrated to his or her asymptotic level of WM span, a decrement in performance relative to the cued condition is related to constraints in processing capacity.1

WM performance on these aforementioned conditions are compared between children with and without RD as a function of their selection of strategies. That is, after stimulus set presentation for the WM tasks, but prior to actual recall, participants are asked to select the strategy they believe will help their retrieval. For the verbal WM task used in this study, the strategy selections are rehearsal, clustering, association, and elaboration. For the visual-spatial WM task, the strategy selections are sectional (focus on recalling sections of a matrix), elemental (focus on key items), global (focus on the gestalt of the task), and backward processing (work backward in reconstructing the patterns) strategies, respectively.2

In summary, Experiment 1 examines three potential, although not mutually exclusive, reasons for the poor WM performance of children with RD. The first determines if strategic knowledge, and/or the stability of strategy choices, underlies poor WM performance. If strategy knowledge contributes to WM performance, then children who perform poorly across such conditions (children with RD) are more likely to select different strategies and/or have less stable strategy choices compared to children with high WM performance (children without RD). The second determines if
processing efficiency is related to poor WM performance. Processing efficiency was defined as the number of cues necessary to achieve an asymptotic level on the gain condition. If differences between reading groups in WM are related to processing efficiency, then comparable span levels between reading groups would emerge for the cued conditions, but the number of cues needed to attain asymptotic performance (gains under the cued condition) between the groups would differ. More specifically, increases in WM span related to the instructional conditions (cuing in this case) would be expected to be greater in children with RD than in children without RD because such procedures would increase the efficiency for retrieving items stored in WM. Thus, a significant Ability Group × Treatment interaction related to WM span would be expected. Finally, the study determines if processing constraints underlie poor WM in children with RD. Although cueing conditions are expected to increase WM performance, it is possible that greater processing constraints emerge for children with RD when they can no longer rely on cues to maintain their asymptotic performance (tasks calibrated to each child’s highest level of performance). Thus, not only would children with RD experience deficits in WM span for the maintenance condition, but the costs of the storage of items would be greater for children with RD than for children without RD. If differences between reading groups in WM performance are related to costs in storage capacity, then procedures that reflect high storage demands (i.e., the maintenance of asymptotic performance after cues have been removed) would be significantly different between the reading groups.

**Method**

Participants. Based on returned parent permission forms, a sample of 45 students (30 boys, 15 girls) was selected from a West Coast public school district. The ethnicity of the sample was 46% Anglo-European, 25% Hispanic, 20% African American, and 9% other (parents of mixed ethnicity). All children with RD were selected from special education classrooms for children with learning disabilities. School district records reported that children had Wechsler IQs in the normal range, but they had word-reading scores below the 25th percentile on either the Wide-Range Achievement Test or the Woodcock Reading Mastery Test. Because we did not have access to these individual test scores, the research staff administered the Test of Reading Comprehension (TORC; Brown, Hammill, & Wiederholt, 1995), the Raven Colored Progressive Matrices Test (Raven, 1976), and the Math subtest from the Woodcock Johnson Psycho-Educational Inventory (WJPI). For this study, students were operationally defined as RD if scores from the Reading subtest of the TORC were below the 25th percentile and math scores on the WJPI were above the 25th percentile. This criterion is consistent with several studies (e.g., Siegel & Ryan, 1989) that use the 25th percentile as cutoff scores for RD. Children without RD yielded comprehension scores greater than the 50th percentile. Twenty-three children formed the RD group (8 girls and 15 boys), and 22 students formed the chronological age matched group (9 girls, 13 boys). No significant differences emerged between the reading groups in terms of gender representation, χ²(1, N = 45) = 1.11, p > .05, ethnicity χ²(5, N = 45) = 74.29, p > .05,
or chronological age, $F < 1$. Also, as shown in Table 1, no significant differences (all $ps > .05$) emerged between the two reading groups on measures of IQ. The mean math score for children with RD was 105.36 ($SD = 13.36$).

WM tasks. The WM tasks in this study required children to hold increasingly complex information in memory while responding to a question about the task. WM tasks typically engage participants in at least two activities after initial encoding: (a) a response to a question or questions about the material or related material to be retrieved and (b) a response to recall item information that increases in set size. The first part of the task is a distracter of initially encoded items, whereas the second part tests storage.

Two tasks from a battery of 11 from a standard measure of WM were selected (the S-Cognitive Processing Test [S-CPT]; Swanson, 1995): Digit-Sentence Span and Mapping/Directions. These tasks were chosen to represent verbal and visual-spatial processing, and each included a delayed recall condition during which a recall strategy was selected. An earlier study (Swanson, 1996) established the construct validity and reliability of the measures. The Cronbach’s alpha for each task used in this study, with age partialed out, was greater than .80 (Swanson, 1995). The current sample reliability for the measures varied from .54 to .96 across the WM conditions.

The purpose of the Digit-Sentence Span task was to assess the participant’s ability to remember numerical information embedded in a short sentence. Before stimulus presentation, the participant was shown a figure depicting four strategies for recalling numerical information. Specifically, the participant was shown procedures on how to remember the address (e.g., 2-4-6-3 Bader Street) with the following instructions: (a) saying the number over to oneself, (b) saying the numbers in pairs, (c) remembering that the numbers 2-4-6-3 go with a particular street, or (d) thinking of other things that go with the numbers. The four pictorial strategies represent rehearsal, chunking, association, and elaboration, respectively. After all strategies had been explained, participants were presented numbers in a sentence context. A sample sentence (Item 3) was, “Now suppose somebody wanted to have you take them to the supermarket at 8 6 5 1 Elm Street. . . .” Numbers were presented 1 every 2 seconds. Children were then presented a process question, “What is the name of the street?” They were then told that they would have to recall the numbers in the sentence in order shortly after they selected from (pointed to) a pictorial array (as shown in Figure 1 in Swanson, 1993b) the strategy that best approximates how he or she will attempt to remember the information. No further information about the strategies shown in the picture was provided to the participant. Participants were allowed 10 seconds in which to point to a strategy.

After a strategy had been chosen, participants were asked to recall the number items in order. If a set of items was recalled correctly, this procedure was repeated with sets of items with increasing difficulty (addresses range from 3 to 14 numbers) until an item was recalled incorrectly. Once an error was made, a series of up to four cues was administered (discussed later), beginning with the least obvious, until the participant recalled the item correctly or failed after receiving all of the cues. For example, for the address 8-6-5-1 Elm Street, the examiner first presented Cue 1, which consisted of the last number (1), followed by Cue 2 (the first number, 8), Cue 3 (the middle numbers, 6 and 5), and on Cue 4, the entire address was repeated to the participant. The participants in each cued condition were told the position of the digit. The range of recall difficulty was 3 digits to 14 digits, and the dependent measure was the number of sets correctly recalled (0 to 9).

The purpose of the Mapping/Directions task was to determine whether the participant could remember a sequence of directions on a map that was void of labels. All students were instructed on strategies prior to testing. As with the Digit-Sentence Span task, presentation phases of the task included stimulus presentation, processing question, strategy selection, and recall. For this task, the experimenter presented the participant with a street map with lines connected to a number of dots that illustrated the direction a bike would go to get through the city (see Figure 2 in Swanson, 1993b). The dots represented stoplights, and the lines and arrows indicated the direction the bicycle should go. Participants were given 10 seconds to study the map prior to its removal. They then answered a process question (“Were there any dots in the first street column?”) and selected a strategy from an array of pictures representing four strategies. The pictures reflected the following memory strategies: (a) begin by filling in the dots or stoplights first and then draw the lines, (b) start with the design and then fill in the dots, (c) do the parts of the city that you remember and then try to put together the rest, or (d) start from the place where they drive (ride) out of the city and then work backward. These four strategies corresponded to elemental, global, sectional, and backward processing of patterns, respectively. Participants were given 10 seconds to choose the strategy that represented how they would remember the map. Children were then given a blank map and were asked to provide the dots (stoplights) and lines (showing where the bike was to go).

If an error on the map was made, and prior to the presentation of each necessary cue, participants were presented with a blank map. In Cue 1, the experimenter drew in the last column (or columns) of dots, lines, and arrows and asked the child to complete the map. For Cue 2, the beginning of the route was filled in on the blank map, and Cue 3 provided the child with only the middle portion of the map. Cue 4 began with the experimenter filling in the entire map and then allowing the participant 10 seconds to study it before providing the participant with another blank map to complete. The maps to be recalled increased in difficulty from
4 to 19 dots. The dependent measure was the number of map sets drawn correctly (range of 0 to 9).

**Span Scores and WM Conditions**

Span scores were calculated for initial, cued, and maintenance conditions. In this study, gain or cued scores were used interchangeably. Span scores for the cued or gain conditions were defined as the highest score that was obtainable with cues. Span scores for the maintenance condition were established by readingministering the highest set of items from the cued condition but without cues. Span scores for the maintenance condition were established after the two WM tasks under initial and cued conditions were administered. The span scores for the maintenance condition reflected the stability of the asymptotic level after the probing condition had been removed. This measure was scored dichotomously so that if the span score for the gain condition was not maintained, the initial score was assigned to the participant. If the span score achieved during the cued conditions was maintained, the span score for the maintenance conditions matched the gain score. A cue or probe score was also computed and was defined as the number of cues (probes) needed to achieve the highest span score under the gain conditions.

**Procedure**

Each child was tested individually. All items for the initial condition were administered until (a) a process question was missed or (b) an error in retrieval occurred. If an error in retrieval occurred (a participant omitted, inserted, or incorrectly ordered the numbers or dots, related to the appropriate task), cues were administered. The only stipulation for beginning the probing during the cued condition was that the process question be answered correctly. Cues were administered based on the type of error made (i.e., whether the error was related to recency, primacy, or middle items), and probing procedures continued until targeted items could not be recalled correctly. After two WM tasks were administered under initial and gain or cued conditions, the examiner readministered the same items for the highest successful set (highest item established under gain conditions) for each task. The general instructions for these maintenance conditions were, “These items were presented to you earlier. I want to see what you can remember this time without hints.”

**Results**

**Ability group comparisons.** A 2 (ability group: RD vs. non-RD) × 2 (modality: verbal vs. visual-spatial WM) × 3 (condition: initial vs. gain vs. maintenance) MANOVA with repeated measures on the last two factors was computed on span scores. Because of inequalities in scaling between the verbal and visual-spatial WM tasks, span scores were converted to z scores for within-task conditions for the total sample. The z scores for the maintenance and gain measurements were based on the mean and standard deviations of the initial condition for verbal (M = 1.84, SD = 1.22) and visual-spatial (M = 1.81, SD = 1.15) WM measures. Although such a procedure eliminated significant main effects related to modality (verbal vs. visual-spatial WM), the procedure allowed for a determination of interactions related to ability group. The z scores for the number of cues needed to establish the span score during the gain condition were also computed for the verbal WM (M = 4.20, SD = 2.06) and visual-spatial (M = 3.02, SD = 2.49) tasks.

A significant main effect was found for ability group, Wilks’s Λ = .53, F(6, 38) = 5.58, p < .001, η² = .47, and condition, Wilks’s Λ = .33, F(6, 42) = 41.23, p < .001, η² = .67. No other significant effects emerged, all ps > .10. More important, no significant Ability Group × Modality interaction occurred, Wilks’s Λ = .98, F(1, 43) = 0.82, p > .05, η² = .02, suggesting that the pattern of results for the two groups was statistically comparable across verbal and visual-spatial domains. In general, the results indicated that children with RD had lower span scores (z score mean across conditions, M = 0.19, SD = 0.47) than did children without RD (M = 0.96, SD = 0.63) across all conditions. A Tukey test showed that mean z scores for WM span were significantly (ps < .05) lower in the initial (M = −0.05, SD = 0.75) than in the gain (M = 1.08, SD = 0.86) and maintenance conditions (M = 0.69, SD = 0.98), initial < maintenance < gain. As shown in Table 1, the univariate were not significant for the number of cues necessary to raise the span level in the gain condition, suggesting that the number of cues to establish asymptotic (gain) performance was comparable between the two reading groups.

Effect sizes were computed by comparing span scores from the gain and maintenance conditions with span scores from the initial condition. Because there was no interaction related to modality, effect sizes were calculated across verbal and visual-spatial conditions. For children with RD, the mean effect sizes (Cohen’s d) were 1.27 and 0.80 for the gain and maintenance conditions, respectively. A similar pattern emerged for children without RD. The mean effect size for children without RD was 1.26 for the gain condition and 0.90 for the maintenance condition. Thus, based on Cohen’s (1988) criterion for high effect sizes (absolute effect sizes > 0.80), both reading groups improved substantially from the initial testing conditions.

**Strategy subgroups.** To compare the effect of individual strategy choices on WM performance, children were subgrouped by the frequency of strategy choices. Strategy
choices were scored according to the most frequent (mode) strategy response on each task. An unstable strategy choice indicated either that (a) all strategies were chosen equally often and/or (b) no strategy was selected within the given time period. No significant differences emerged in strategy choices between children with or without RD on the verbal WM task, $\chi^2(4, N = 45) = 3.42, p > .05$, or the visual-spatial WM task, $\chi^2(4, N = 45) = 4.92, p > .05$. Thus, although some children had more opportunities to make strategy choices because of higher WM performance, no differences were found between reading groups in the frequency of choices. Table 2 shows the percentage of choices for each strategy as a function of reading group.

As shown in Table 2, there was a trend showing that more unstable strategy choices emerged for children with RD than for children without RD. A comparison was made between children whose strategy choices were unstable and those whose strategy choices were stable. Because the Phi coefficient between strategy stability (stable vs. unstable choices) and modality (visual-spatial vs. verbal WM measure) was not significant, $\phi = -.17, p > .05$, children were classified as unstable strategy choosers if an unstable strategy choice occurred for the either verbal or visual-spatial WM task. The remaining children were designated as stable strategy users. Significant differences emerged between children with and without RD as a function of strategy stability (stable vs. unstable strategy choices), $\chi^2(1, N = 45) = 3.94, p < .05$. Six children (27%) without RD made unstable choices, whereas 13 children (56%) with RD made unstable strategy choices for either the verbal and/or visual spatial task. Because no significant interaction effects emerged related to modality (verbal vs. visual-spatial WM) and reading group on the span scores, children who yielded stable versus unstable strategy choices were compared on composite $z$ scores related to the initial, gain, and maintenance conditions.

A MANOVA was performed to determine if the effect of strategy choice on WM performance across the initial, gain, and maintenance conditions was significant. A significant main effect was found for stability group, Wilks’s $\Lambda = .82$, $F(3, 41) = 2.80, p < .05, \eta^2 = .18$, and condition, Wilks’s $\Lambda = .37, F(2, 42) = 37.53, p < .001, \eta^2 = .67$. No other significant effects emerged, $p > .05$. The results indicated that stable strategy choosers had higher WM $z$ scores ($M = 0.76, SD = 0.72$) than did unstable strategy choosers ($M = 0.31, SD = 0.50$). A Tukey test indicated that the $z$ scores for WM span were significantly ($p < .05$) higher for stable strategy choosers for the gain ($M = 1.33, SD = 0.74$) than for the maintenance ($M = 0.82, SD = 0.84$) condition, and WM span scores for the maintenance condition were higher than for the initial condition ($M = 0.14, SD = 0.73$). Likewise, for children who made unstable strategy choices, the $z$ scores were higher for span scores in the gain ($M = 0.74, SD = .58$) than in the maintenance ($M = 0.51, SD = 0.67$) condition, and span scores were higher in the maintenance than in the initial condition ($M = -0.31, SD = 0.71$).

In general, the important findings were that (a) children with RD yielded lower span scores than did children without RD across all conditions, (b) effect sizes related to span scores for the gain and maintenance conditions were comparable between reading groups, (c) unstable strategy choices were more likely to occur for children with RD than for children without RD, and (d) children who made unstable strategy choices yielded lower WM span scores than did those who made stable strategy choices.

### Processing Efficiency

Although strategy instability may account for low WM performance in children with RD, it may also be the case that their poor WM performance was due to limitations in processing efficiency. This scenario seems unlikely, however, because both groups were statistically comparable in the number of cues necessary to reach their highest span levels in the gain condition. However, it is possible that although the use of cues did not separate ability groups, differences in span scores between memory conditions may provide a more direct assessment of processing efficiency. To consider this possibility, a one-way ANOVA was computed on process efficiency scores. Process efficiency scores were calculated as: [(maintenance – initial span scores)/number of probes]. A composite $z$ score across the two WM tasks was computed. Scores close to zero were reflective of high processing efficiency. The ANOVA showed no significant group effect, $F(1, 43) = 0.17, p > .05, \eta^2 = .003$. Processing

### Table 2. Strategy Choices as a Function of Reading Group in Experiment I

<table>
<thead>
<tr>
<th>Strategy Choice</th>
<th>Children With Reading Disabilities</th>
<th>Children Without Reading Disabilities</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$n$</td>
<td>$%$</td>
</tr>
<tr>
<td>Verbal working memory</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unstable</td>
<td>9</td>
<td>39</td>
</tr>
<tr>
<td>Rehearsal</td>
<td>3</td>
<td>13</td>
</tr>
<tr>
<td>Clustering</td>
<td>5</td>
<td>22</td>
</tr>
<tr>
<td>Association</td>
<td>5</td>
<td>21</td>
</tr>
<tr>
<td>Elaboration</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>Visual-spatial working memory</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unstable</td>
<td>7</td>
<td>30</td>
</tr>
<tr>
<td>Elemental</td>
<td>6</td>
<td>22</td>
</tr>
<tr>
<td>Global</td>
<td>7</td>
<td>30</td>
</tr>
<tr>
<td>Sectional</td>
<td>3</td>
<td>12</td>
</tr>
<tr>
<td>Forward</td>
<td>1</td>
<td>4</td>
</tr>
</tbody>
</table>
The results were again nonsignificant, \( p < .01 \). The processing efficiency that included span scores for the gain condition \( (\eta^2 = 0.005) \). Thus, the results do not support the hypothesis that reading group differences in WM can be attributed to limitations in processing efficiency.

### Demands on Storage Capacity

Span scores from the maintenance condition reflected greater demands on storage capacity than those from the initial and gain conditions. However, it was unclear whether greater demands were placed on WM capacity in children with RD when compared to children without RD. To address this issue, we adapted a formula from Cowan (2001) and Baddeley, Sala, Papagno, and Spinnler (1997; MU score) to assess the performance costs on capacity. The formula takes into consideration the highest number that could be remembered (span scores related to the cued or gain condition, in this case) for each participant. Costs in capacity in this case were calculated as the performance cost for WM span from the maintenance and initial conditions, capacity costs = gain (cued score) \( \times \) (maintenance score – initial score). The span score from the gain or cued conditions was the highest set size established by the participant, and therefore, this score was used to calibrate performance. That is, costs or constraints on a limited capacity system take into consideration individual differences in the highest span score achieved. For example, using the above formula, a skilled reader with a z score of 0.50, 0.35, and 0.35 for gain, maintenance, and initial scores, respectively, would yield the same performance cost score (0) as a child with RD receiving a z score of -0.50, -0.35, and -0.35, respectively. Lower scores for children with RD when compared to children without RD reflected greater demands or constraints on a limited capacity system.

The ANOVA showed a significant group effect on measures of capacity, \( F(1, 43) = 7.54, p < .01, \eta^2 = .15 \). Significantly higher scores (reflecting less demands on capacity) emerged for children without RD \( (M = 0.30, SD = 0.27) \) when compared to children with RD \( (M = 0.17, SD = 0.13) \). The results suggested that capacity demands were greater for children with RD than for children without RD.

**Correlations.** Our next analysis determined whether unstable strategy choices, strategy efficiency, and/or capacity were significantly correlated with reading comprehension. The intercorrelations among the WM, reading comprehension, and strategy dummy variables (coded as 0 and 1, for stable and unstable, respectively; processing efficiency \( [\text{processing efficiency} = (\text{gain} – \text{initial span score})/\text{probes}] \); and capacity are shown in Table 3. The intercorrelations yielded three important findings. First, the stability variable was significantly related to span scores under the gain and initial conditions but not to scores for the maintenance condition. Thus, strategy stability was related to conditions that increased WM performance when compared to conditions that placed high demands on capacity. However, the magnitude of these correlations was small. (Note, however, that when controlling for an inflation for alpha, .05/36, \( p < .001 \), none of the correlations related to strategy stability met the critical alpha level. However, because of the sample size and to control for Type II error, the alpha was set at .05.) No other significant effects emerged for stability scores. Second, the number of cues or probes administered was significantly correlated with gain scores but not with other variables. Finally, reading comprehension performance was significantly correlated with all WM span scores and with the capacity score.

**Hierarchical regression.** The final analysis determined whether strategy stability, processing efficiency, and/or processing capacity contributed unique variance to reading comprehension performance. To investigate this relationship, three hierarchical models were tested. Stability scores, efficiency scores, or capacity scores were first entered in the separate models (Models 1a, 1b, and 1c, respectively).

### Table 3. Intercorrelations Between Achievement, Working Memory, and Strategy Choice for Experiment 1

<table>
<thead>
<tr>
<th>Variables</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. TORC</td>
<td>–</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Stability</td>
<td>-.29</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Initial</td>
<td>.56***</td>
<td>-.30*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. Gain</td>
<td>.48***</td>
<td>-.34*</td>
<td>.41**</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. Maintenance</td>
<td>.63***</td>
<td>-.20</td>
<td>.59***</td>
<td>.71***</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. Cues</td>
<td>.01</td>
<td>-.04</td>
<td>-.14</td>
<td>.46**</td>
<td>.13</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. Capacity</td>
<td>.45**</td>
<td>-.15</td>
<td>.11</td>
<td>.62***</td>
<td>.75***</td>
<td>.37*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8. Efficiency</td>
<td>.24</td>
<td>.02</td>
<td>.14</td>
<td>.21</td>
<td>.22</td>
<td>.21</td>
<td>.33*</td>
<td></td>
</tr>
</tbody>
</table>

Note: TORC = Test of Reading Comprehension.

*p < .05, **p < .01, ***p < .001.
In Models 2 and 3, the regression model simultaneously entered strategy stability and processing efficiency scores (Model 2) followed by the capacity score (Model 3).

As shown in Table 4, reading comprehension was significantly predicted by strategy stability and capacity in Models 1a and 1c, respectively. The results showed that significant variance related to predicting comprehension was related to strategy stability (8% of the variance) and capacity (20% of the variance) scores. In Model 2, the results showed that strategy stability, but not processing efficiency, contributed significant variance to reading performance. Model 2 was significant, $R^2 = .14$, $F(2, 42) = 3.62$, $p < .01$, but the increment in $R^2$ (6%) when compared to Model 1a was not significant, $F_{inc}(1, 42) = 3.00$, $p > .05$.

As shown in Model 3, entry of the capacity score contributed unique variance in predicting comprehension performance ($p < .05$). Two important findings emerged in Model 3. First, strategy stability was not significant in predicting comprehension scores. Thus, it appeared that consistency in strategy knowledge choices was moderated by processing capacity. The incremental improvement of $R^2$ from Model 2 to Model 3 was significant for predicting comprehension scores, $F_{inc}(1, 41) = 7.64$, $p < .001$.

| Table 4. Hierarchical Regression Model on Reading Comprehension Scores for Experiment 1 |
|-----------------|---------|--------|-----|--------|
|                  | B       | SE     | $\beta$ | t ratio |
| **Model 1a**    |         |        |      |        |
| Strategy stability | -.29    | .14    | -.29 | 1.99** |
| $R^2$            | .08     |        |      |        |
| $F$ ratio (1, 43) | 3.95**  |        |      |        |
| **Model 1b: Efficiency** |         |        |      |        |
| Processing efficiency | .24     | .14    | .25  | 1.64   |
| $R^2$            | .06     |        |      |        |
| $F$ ratio (1, 43) | 2.69    |        |      |        |
| **Model 1c: Capacity** |         |        |      |        |
| Processing capacity | .45     | .13    | .45  | 3.30** |
| $R^2$            | .20     |        |      |        |
| $F$ ratio (1, 43) | 10.91** |        |      |        |
| **Model 2: Strategy stability and efficiency** |         |        |      |        |
| Strategy stability | -.29    | .14    | -.30 | -2.09* |
| Efficiency       | .25     | .14    | .25  | 1.76   |
| $R^2$            | .14     |        |      |        |
| $F$ ratio (2, 42) | 3.62*   |        |      |        |
| **Model 3: Strategy, efficiency, and capacity** |         |        |      |        |
| Strategy stability | -.23    | .13    | -.23 | -1.75  |
| Efficiency       | .37     | .14    | .37  | 2.59** |
| Capacity         | .27     |        |      |        |
| $F$ ratio        | 4.97**  |        |      |        |

*p < .05, **p < .01.

Overall, the important results were that demands on processing capacity rather than on strategy stability and/or processing efficiency were a reliable predictor of reading comprehension performance.

**Summary**

Three important results emerged. First, although skilled readers outperformed children with RD across WM condition, effect sizes related to WM improvement were comparable between reading groups on gain and maintenance conditions. Second, no differences emerged between reading groups in knowledge of strategies assumed to improve WM performance. However, stable strategy choices, rather than unstable choices, were related to high and low WM span performance. Finally, the results suggested that the locus of group differences in reading comprehension were best predicted by measures that show demands on WM capacity rather than by measures of strategy stability or processing efficiency.

There are at least two limitations to Experiment 1 that may account for the moderate effects related to strategy stability. The most critical is that only declarative knowledge was compared between the two ability groups. Declarative knowledge was not linked to procedural knowledge. That is, children were asked to choose from a menu the strategies that they believed would best help them retrieve previously presented information, but there was no indication of the extent to which the strategies selected were actually employed. It may be that the measures of strategy selection provided in these tasks are not fine-tuned enough to access children’s understanding of the use of a systematic approach to memory tasks.

Second, although Experiment 1 showed that the WM span of children with RD can be significantly improved upon, there was minimal evidence that such procedures reduced the variance between ability groups. In fact, the mean effect sizes across the conditions were comparable between the ability groups. That is, a key assumption in assessing the positive effects of strategy training studies is that the variance between the two groups will be reduced. Although this occurred somewhat for the visual WM task (the standard deviations were smaller in the gain condition than in the initial condition), the results overall are best captured by differences in capacity. Thus, the purpose of this next experiment was to directly train children in strategy use in order to reduce the variance between the groups.

**Experiment 2**

Experiment 2 addressed the question as to whether direct strategy instruction can reduce the variance in WM
performance between children with and without RD. Although Experiment 1 showed that WM span can be significantly improved upon under cued recall conditions, the results also suggested that residual differences between the two reading groups were more likely due to constraints in WM capacity rather than to variations in strategy choices. For Experiment 2, groups were randomly assigned clinical trials that involved rehearsal training. Rehearsal training was selected as the instructional condition because Turley-Ames and Whitfield (2003) found that WM span scores increased for average-achieving adults as a result of using rehearsal strategy relative to other strategies (e.g., clustering, imagery).

The purpose of Experiment 2 was to determine if rehearsal training improves the WM performance of children with RD relative to children without RD. Of critical interest, however, is whether the training effects are greater in children with RD than in children without RD. The finding is of interest because it is assumed that strategy training helps allocate WM resources more efficiently in low-span participants (see Turley-Ames & Whitfield, 2003, for further discussion of this hypothesis). Another issue of concern is whether strategy training reduces the variance between WM and reading. If this is so, then the magnitude of the correlation between WM and achievement should be smaller at posttest than at pretest. The competing hypothesis, however, is that although strategy training may provide greater gains in performance for children with RD, such training does not account for the correlations between WM span and reading. That is, something else drives the predictive power of WM measures. A final issue of concern was to determine what underlies the strategy training effects on the WM performance of children with RD. There are at least three sources of individual differences in children’s WM performance that strategy training may influence. One relates to storage, another to processing, and another to the interchange between processing and storage on WM tasks (see Jarrold & Bayliss, 2007). To account for each of these mechanisms, measures were taken of the correct responses to item (storage) and process questions as well as the trade-off between variations in storage and processing. Based on the results of Experiment 1, it was assumed that the source of individual differences in WM performance as a function of strategy conditions would be isolated to demands on storage.

**Method**

**Participants.** Participants were 57 fifth- and sixth-grade students from a low-income elementary school from a large urban Southern California school district. Consent letters were distributed to 336 students and their parents, and 102 were returned. These children were tested on the Raven Colored Progressive Matrices Test. Children who scored below a standard score of 90 (25th percentile) were not selected for further testing. Children who scored above a standard score of 90 were administered the Reading and Math subtests of the Wide Range Achievement Test (WRAT-III; Wilkinson, 1993). Participants who scored at or below the 25th percentile on the Reading subtest (n = 29) of the WRAT-III were operationally defined as RD. These children were currently spending a least 1 hour in a resource room for children with learning disabilities. Participants in the regular classroom who scored above the 50th percentile on the WRAT-III Reading subtest (n = 29) were operationally defined as average readers. Thus, of the 102 children screened for inclusion, 58 children met the research criteria.

Participants from each reading group were randomly assigned to a control or treatment condition (discussed below). The percentage of ethnic representation was 38% Anglo, 51% Hispanic, 4% Black, and 7% other. For ethnicity, no significant differences emerged between reading groups; χ²(3, N = 58) = 4.16, p > .05. The gender representation was 46% female and 54% male. For gender presentation, no significant differences emerged between reading groups, χ²(1, N = 58) = .007, p > .05. All participants were native English speakers.

**Tasks**

**Operation Span.** A version of the Turley-Ames and Whitfield (2003) Operation Span task was modified for children in this study. Previous studies have shown a significant relationship between the Operation Span task and academic achievement (Siegel & Ryan, 1989; Tows, Hitch, & Hutton, 2001). The Operation Span task was also selected because the processing activity (math computation; see below) drew up intact skills for children with RD (i.e., their math scores were in the average range). Math problems and to-be-remembered (TBR) words were displayed in black on white 5××-7-inch cards. The Operation Span task (shown in the appendix) assessed WM span by having participants solve simple math problems while remembering unrelated TBR words that follow each math problem. After each simple addition or subtraction operation, a TBR word was visually presented for later recall. Our measure differed from those of Turley-Ames and Whitfield in the following two ways. First, a list of high-frequency words derived from the Fry Most Frequently Used Words and Dolch reading lists served as the TBR words for pre- and post-Operation Span measures. Second, only one-digit addition and subtraction math problems were used. Prior to the study, TBR words were assigned randomly to math operations. All children were pretested on their knowledge of these words. All words
were below children’s reading level by two grades. Similar to Turley-Ames and Whitfield’s measure, operation word sequences were presented in five parts: (a) a number from 1 to 18, (b) an addition or subtraction sign, (c) a number from 1 to 18, and (d) “ = ____.” When the “d” part of the operation was presented, the participant read the math problem aloud and reported an answer, and the experimenter recorded the participant’s answer. After providing an answer for the math problem, the TBR word was revealed for 5 seconds and read aloud by the participant.

Operation word sequences were presented in increasing set size. Children completed two practice trials with a set size of two. Children were then presented with operation word sequences in sets of 2, 3, 4, and 5, with two trials for each set size, for a total of 10 sets. To ensure that the results were not due to order effects, two versions of test stimuli (Form A and Form B) were assigned randomly within experimental trials. Children received points toward their span scores for correctly solving the math problems, for the number of correctly recalled words, and for correct order of word recall. This scoring procedure was implemented to prevent giving participants credit for recalling words at the expense of solving the math problems incorrectly.

Because children may differ in the extent to which they distribute attention between the process (math problems) and the storage (words), it was desirable to consider these components in scoring. Although we report item correct scores (see Table 5), for measuring treatment effects at pretest and posttest we adapted a formula from dual-task studies (see Baddeley et al., 1997; MU score) to assess the influence of treatment performance costs. Performance costs were calculated using a formula that included the processing scoring (total processing question correct/total possible) minus the storage score (proportion of items correctly recalled [total correct items/total possible]) divided by the proportion of processing questions correct plus the storage scores (processing – storage/processing + storage). A higher score reflects more cost (higher trade-off related to storage), whereas a lower score indicates less performance trade-off between storage and processing. Both pre- and post-Operation Span tests were scored in this manner.

Transfer measure. To check for transfer-of-strategy instruction effects, pretest and posttest performance of children’s adaptations (Swanson, 1992) of Daneman and Carpenter’s (1980) Listening Sentence Span task was assessed. The Listening Sentence Span task required the presentation of groups of sentences, read aloud, for which children tried to simultaneously understand the sentence contents and to remember the last word of each sentence. The number of sentences in the group gradually increased from 2 to 6. After each group was presented, the participant answered a question about a sentence and then was asked to recall the last word of each sentence. The dependent measure was the total number of correctly recalled word items up to the largest set of items (e.g., Set 1 contained 2 items, Set 2 contained 3 items, Set 3 contained 4 items, etc.) in which the process question was also answered correctly.

State-mandated districtwide tests. Reading comprehension and math computation scores from the California Standardized Testing and Reporting (STAR) program were used in analysis. This standardized test is administered in the majority of California schools and was mandated by the state legislature in 1997. Raw scores from the portion of the STAR that assessed reading comprehension and mathematics were used in analysis.

Treatment Conditions

For both treatment and control conditions, during the pretest children were encouraged to work as quickly and accurately as possible through the test stimuli of the Operation Span task. Children were required to read aloud each part of the operation word sequence as it appeared. This included each part of the math problems, the answer to the math problems, and the TBR words. After each set, the experimenter said, “Recall words,” and participants attempted a free recall of the TBR words from the preceding set. Children were given as much time as necessary to recall words. They were also encouraged to guess if they could not recall a word.

After completing the Operation Span pretest, half of the children with RD and half of the children without RD were provided rehearsal strategy instructions. Children receiving rehearsal strategy instruction were told,

Before starting a second version of the task you just completed, I would like you to try a particular strategy that may improve your performance. As before, work as quickly and as accurately as possible on the math operations. When you are presented with a to-be-remembered word, I would like you to say that word aloud as many times as you can before going on to the next math problem. As more words are added to the set, please say aloud not only the new word, but also other words presented previously in the set. In other words, each time you are presented with a new to-be-remembered word, say that word aloud and any previous to-be-remembered words in that set as many times as you can.

To ensure children were using the described rehearsal strategy, participants were asked to perform the rehearsal task aloud. The procedure was modeled during the practice trials, and children were reminded, if necessary, to use the
strategy during the subsequent trials. If a child forgot to repeat previously presented words, he or she was prompted with the phrase, “previous words.” All children complied with this procedure.

Children randomly assigned to the control condition were given a paraphrased version of the original instructions before completing the Operation Span posttest. Prior to the administration of the posttest, all children completed three additional practice trials. The approximate time for pretests, treatment, and posttests for each child varied from 140 to 160 minutes and occurred over 3 days (Sessions 1, 2, and 3, respectively). Pretest occurred in Session 1, treatment in Session 2, and posttest in Session 3. The time devoted to strategy instruction and practice in the treatment session varied from 10 to 15 minutes.

In both of the experiments and consistent with the procedures outlined by Turley-Ames and Whitfield (2003), the presentation of words during training were experiment-paced. This was done to provide some controls for study time of the words between the two conditions.

Procedures. All tests were administered individually by a graduate student in a quiet classroom at the child’s school. All participants were administered the Raven Colored Progressive Matrices Test and the WRAT-III Reading and Math subtests. Participants were assigned randomly to one of the two versions of the pre- and posttest measures. The order of administration of the WM test (Listening Sentence Span, Operation Span) also included two versions randomly assigned to pretest and posttest conditions.

Results

Means and standard deviations for age, scores for the Raven Colored Progressive Matrices Test, WRAT-III Reading and Math subtests, all WM pre- and posttests, and California Standards (CST) Reading and Math composite tests as a function of ability group and treatment conditions are displayed in Table 5. The cell size for participants administered the CST was reduced because these scores were not available for all the students in the sample. In addition, 3 children

---

Table 5. Means and Standard Deviations for Age and Scores on All Measures for Experiment 2

<table>
<thead>
<tr>
<th>Classification</th>
<th>Control (n = 14)</th>
<th>Treatment (n = 15)</th>
<th>Control (n = 14)</th>
<th>Treatment (n = 15)</th>
<th>Reliability</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>M (SD)</td>
<td>M (SD)</td>
<td>M (SD)</td>
<td>M (SD)</td>
<td></td>
</tr>
<tr>
<td>Raven (percentile)</td>
<td>11.41 (0.87)</td>
<td>11.07 (0.91)</td>
<td>11.10 (0.63)</td>
<td>11.43 (0.64)</td>
<td></td>
</tr>
<tr>
<td>WRAT-III Reading (percentile)</td>
<td>55.14 (17.49)</td>
<td>57.33 (19.72)</td>
<td>62.21 (20.66)</td>
<td>49.6 (24.05)</td>
<td>.74</td>
</tr>
<tr>
<td>WRAT-III Math (percentile)</td>
<td>58.43 (19.08)</td>
<td>52.94 (16.98)</td>
<td>15.86 (7.42)</td>
<td>12.87 (7.91)</td>
<td>.75</td>
</tr>
<tr>
<td></td>
<td>75.29 (24.79)</td>
<td>76.61 (25.05)</td>
<td>51.86 (27.9)</td>
<td>42.27 (32.47)</td>
<td>.85</td>
</tr>
</tbody>
</table>

| Transfer measure        |                  |                    |                  |                    |             |
| Listening span          |                  |                    |                  |                    |             |
| Pretest item            | 11.57 (3.08)     | 11.40 (2.93)       | 9.64 (4.03)      | 10.27 (3.31)       | .76         |
| Posttest item           | 13.43 (1.74)     | 14.53 (1.80)       | 10.71 (3.41)     | 11.27 (3.84)       | .78         |
| Pretest process question| 1.37 (.84)       | 1.66 (1.17)        | 1.00 (.67)       | 1.26 (.79)         | .51         |
| Posttest process question| 2.00 (1.03) | 2.00 (0.75)      | 1.71 (0.99)      | 1.40 (1.35)        | .43         |
| Pretest cost            | -.038 (.29)      | -.32 (.36)         | -.45 (.33)       | -.36 (.30)         |             |
| Posttest cost           | -.27 (.24)       | -.28 (.19)         | -.26 (.34)       | -.46 (.42)         |             |

| Training measure        |                  |                    |                  |                    |             |
| Operation span          |                  |                    |                  |                    |             |
| Pretest item            | 27.36 (3.77)     | 25.60 (3.56)       | 19.93 (5.47)     | 17.53 (5.88)       | .86         |
| Posttest item           | 25.57 (4.64)     | 33.22 (2.65)       | 20.64 (4.27)     | 28.93 (5.68)       | .90         |
| Pretest process question| 27.57 (0.85)     | 27.86 (0.35)       | 27.21 (0.80)     | 26.26 (.228)       | .62         |
| Posttest process question| 27.21 (1.31) | 27.86 (0.51)      | 27.42 (0.64)     | 24.86 (3.41)       | .84         |
| Pretest cost            | 0.08 (0.06)      | 0.05 (0.07)        | 0.17 (0.13)      | 0.21 (0.15)        |             |
| Posttest cost           | 0.03 (0.09)      | -.08 (0.04)        | 0.14 (0.10)      | -.06 (0.07)        |             |

| Group test              |                  |                    |                  |                    |             |
| CST Reading             | 53.36 (12.56)    | 47.08 (12.75)      | 47.23 (23.13)    | 36.21 (19.15)      | .95         |
| CST Math                | 41.21 (14.44)    | 42.92 (9.89)       | 37.85 (21.18)    | 37.14 (23.87)      | .96         |

Note: Raven = Raven Colored Progressive Matrices; WRAT-III = Wide Range Achievement Test; item = item recall for storage component of working memory; CST = California Standards Test.
(2 in control and 1 in treatment) participated in the pretest but not the posttest. Thus, the analysis of treatment included only the analysis of complete cells. Reliability was calculated on WM measures using Cronbach’s alpha. Reliability on all experimental measures was between .76 and .90 for the storage component and between .43 to .84 for the processing component of WM (see Table 5). To ensure variance in our analysis of WM task performance, we used the number of items correctly recalled rather than absolute span as the dependent measure (see Friedman & Miyake, 2004, and Unsworth & Engle, 2007, p. 1047, for advantages of scoring items correct across sets vs. absolute scoring, which is highest set recalled perfectly).

**Classification.** Prior to the analysis, the ability groups were compared on the classification measures. In terms of chronological age, no significant differences emerged between groups, $F(1, 57) = 0.04, p > .05$; treatment condition, $F(1, 54) = 0.01, p > .05$; or Group × Condition interaction, $F(1, 54) = 2.71, p > .05$. Performance on the Raven test yielded no significant differences as a function of ability group, $F(1, 54) = 0.26, p > .05$; condition, $F(1, 54) = 1.02, p > .05$; or Group × Condition interaction, $F(1, 54) = 2.60, p > .05$.

As expected, significant differences between the reading groups emerged for WRAT-III reading scores in favor of the children without RD, $F(1, 54) = 98.82, p < .001$. However, no significant differences in reading scores were related to treatment condition, $F(1, 54) = 1.05, p > .05$, or to Group × Condition interaction, $F(1, 54) = 0.23, p > .05$.

For math scores, significant reading group differences emerged in favor of children without RD, $F(1, 54) = 18.19, p < .001$. No significant differences emerged as a function of treatment condition, $F(1, 54) = 0.15, p > .05$, or Group × Condition interaction, $F(1, 54) = 0.40, p > .05$.

**Pretest performance.** Prior to comparing treatment effects, a MANOVA was computed on pretest WM item (storage) scores for the Operation Span and Listening Sentence Span tasks as a function of ability group and treatment. As expected, the MANOVA was significant for ability group, Wilks’s $\Lambda = .58, F(2, 53) = 18.69, p < .001, \eta^2 = .42$. However, no significant differences emerged as a function of treatment, Wilks’s $\Lambda = .94, F(2, 53) = 1.55, p > .05, \eta^2 = .06$, or the Ability Group × Treatment interaction, Wilks’s $\Lambda = .99, F(2, 53) = 0.17, p > .05, \eta^2 = .01$.

A similar analysis was computed on accuracy scores related to the process questions. The MANOVA was significant for ability group, Wilks’s $\Lambda = .83, F(2, 53) = 5.28, p < .001, \eta^2 = .17$. However, no significant differences emerged as a function of treatment, Wilks’s $\Lambda = .96, F(2, 53) = 1.16, p > .05, \eta^2 = .04$, or the Ability Group × Treatment interaction, Wilks’s $\Lambda = .94, F(2, 53) = 1.61, p > .05, \eta^2 = .07$.

Because no significant differences in pretest performance emerged as a function of treatment conditions, both pretest performance and posttest performance were considered as repeated measures in the subsequent analysis.

**Treatment outcomes.** A 2 (ability group: RD vs. non-RD) × 2 (treatment condition: rehearsal vs. control) MANOVA, with repeated measures on the last factor, was computed on items correctly recalled on the Operation Span task. A significant main effect emerged for group, Wilks’s $\Lambda = .58, F(1, 54) = 18.75, p < .001, \eta^2 = .42$; treatment condition, Wilks’s $\Lambda = .36, F(1, 54) = 45.22, p < .001, \eta^2 = .64$; and testing wave, Wilks’s $\Lambda = .45, F(1, 54) = 63.84, p < .001, \eta^2 = .55$. No significant Ability Group × Treatment interaction emerged, $F < 1.0$. Thus, no support was found for the assumption that treatment conditions reduce the variance between reading groups. Significant interactions emerged for Ability Group × Test Wave, Wilks’s $\Lambda = .87, F(1, 54) = 7.88, p < .05, \eta^2 = .13$, and the Treatment Condition × Test Wave, Wilks’s $\Lambda = .40, F(1, 54) = 80.01, p < .001, \eta^2 = .59$. The results showed that children without RD (nondisabled reader [NDR]) outperformed children with RD, performance was better for treatment than for control conditions, and item recall scores were higher at posttest than at pretest. These findings were qualified by the interactions. The Condition × Testing Wave interaction reflected the significant effects for the treatment condition at posttest, $F(1, 54) = 45.96, p < .001, \eta^2 = .39$, but not at pretest, $F(1, 54) = 2.73, p > .05, \eta^2 = .02$.

To investigate the Ability Group × Test Wave interaction, a test of simple effects was performed on difference scores (item recall at posttest – item recall at pretest). Children with RD showed significantly greater gains in item recall than did children without RD (RD $M = 6.24, SD = 6.79$ vs. NDR $M = 3.06, SD = 6.43$), $F(1, 53) = 8.01, p < .05, \eta^2 = .12$. The Ability Group × Test Wave interaction is shown in Figure 1a. To facilitate comparison, pretest and posttest scores were converted to $z$ scores ($M = 0, SD = 1$) based on the total sample. The $z$ scores for the posttest were based on the sample mean and standard deviation at pretest. As shown in Figure 1a, greater gains, regardless of treatment condition, emerged for children with RD than for children without RD.

A similar analysis was computed on the number of correct responses for the process section (computation) of the Operation Span task (see Table 5 for mean scores). A significant main effect emerged for group, Wilks’s $\Lambda = .83, F(1, 54) = 10.83, p < .001, \eta^2 = .27$, and the Group × Condition interaction, Wilks’s $\Lambda = .83, F(1, 54) = 5.17, p < .01, \eta^2 = .16$. No other significant effects emerged, $ps > .05$. The results showed that children without RD yielded higher correct scores ($M = 27.63, SD = 0.76$) than did children with RD ($M = 26.41, SD = 1.97$). These findings were
finding was the Ability Group
The covariate was pretest performance. The only reliable
computed on process questions on the Operation Span task.
(treatment condition: rehearsal vs. control) ANCOV A was
= .32, but not for the control condition,
(1, 27) 
= .08. The results showed that the trade-off scores were more
negative for children with RD than for children without
RD (RD M = -0.16, SD = 0.17 vs. NDR M = -0.05, SD = 0.11) and more negative for the treatment than for the
control condition (treatment M = -0.21, SD = 0.12, vs.
control M = 0.004, SD = 0.08). A test of simple effects
indicated that trade-off scores between ability groups were
significant for the treatment condition (NDR M = -0.13, SD
= 0.08, vs. RD M = -0.29, SD = 0.11), F(1, 28) = 16.86, p < .001,
, η² = .37, but not for the control condition (NDR M =
0.03, SD = 0.08, vs. RD M = -0.02, SD = 0.09), F(1, 26) =
2.44, p > .05, η² = .08. Overall, the results suggested that
demands related to the trade-off between processing and
storage were greater for children with RD than for children
without RD during the treatment conditions.

Transfer-listening span. The next analysis determined if the
treatment effects influenced performance on other WM

qualified by the Group × Condition interaction. As shown
in Table 5, a significant difference was found in favor of
children without RD when compared to children with RD
for the treatment condition, F(1, 28) = 13.44, p < .001, η²
= .32, but not for the control condition, F < 1. These effects
may be due to group differences at pretest. Thus, a 2 (ability
group: children with RD vs. children without RD) × 2
(treatment condition: rehearsal vs. control) ANCOVA was
computed on process questions on the Operation Span task.
The covariate was pretest performance. The only reliable
finding was the Ability Group × Treatment interaction F(1,
53) = 6.73, p < .01, η² = .12. The results showed that there
were significantly more correct responses from children
without RD than from children with RD in the treatment
condition (least square mean [LSM] = 27.47 vs. 25.48),
F(1, 27) = 8.16, p < .01, η² = .12, but not in the control
condition (LSM = 29.99 vs. 27.43), F < 1.

Taken together, the important results were that no support
was found for the assumption that treatment conditions
significantly reduce the variance between reading groups
for the processing or storage of items. Instead, the results
related to processing scores suggest that differences
between reading groups were greater for the treatment
than for the control condition. Thus, as the processing
demands increase, there may be a trade-off with storage
scores. To investigate this possibility, the next analysis
considered group differences in the trade-off (interchange)
between the process and storage components of the
Operation Span task.

Performance costs. The next analysis determined the
extent to which children differed in the distribution of their
attention between the process and storage questions. To
address this issue, we adapted a formula from Baddeley et
al. (1997; MU score) to assess the performance trade-offs
between processing and storage. The child’s trade-off score
was calculated as

\[
\text{Trade-off} = \frac{\text{proportion correct process} - \text{proportion correct storage}}{\text{proportion correct process} + \text{proportion correct storage}}.
\]

The denominator to calculate the proportions was the same
for both storage and processing (56). Table 5 shows the
mean trade-off scores for both pretest and posttest as a
function of treatment and ability group. The dependent
measure for this analysis was the difference in trade-off
scores between pretest and posttest (difference = trade-off
scores at posttest – trade-off scores at pretest) as a function
of ability group and treatment condition. To interpret these
scores, it was assumed that the lower or more negative the
scores relative to the control conditions or comparison
group, the higher the processing demands. A significant
effect emerged for group, Wilks’s Λ = .76, F(1, 54) = 16.97,
p < .001, η² = .24; condition, Wilks’s Λ = .42, F(1, 54) =
73.64, p < .001, η² = .58; and the Ability Group × Condition
interaction, Wilks’s Λ = .92, F(1, 54) = 4.53, p < .05, η²
= .08. The results showed that the trade-off scores were more
negative for children with RD than for children without
RD (RD M = -0.16, SD = 0.17 vs. NDR M = -0.05, SD = 0.11) and more negative for the treatment than for the
control condition (treatment M = -0.21, SD = 0.12, vs.
control M = 0.004, SD = 0.08). A test of simple effects
indicated that trade-off scores between ability groups were
significant for the treatment condition (NDR M = -0.13, SD
= 0.08, vs. RD M = -0.29, SD = 0.11), F(1, 28) = 16.86, p < .001,
, η² = .37, but not for the control condition (NDR M =
0.03, SD = 0.08, vs. RD M = -0.02, SD = 0.09), F(1, 26) =
2.44, p > .05, η² = .08. Overall, the results suggested that
demands related to the trade-off between processing and
storage were greater for children with RD than for children
without RD during the treatment conditions.

Figure 1. Mean z Score Pretest and Posttest Performance as a
Function of Ability Group and Condition: (A) Training Measure –
Operation Span, (B) Transfer Measure – Listening Sentence Span
Note: NDR-C = children without reading disabilities (RD), control group;
NDR-TRT = children without RD, treatment group; RD-C = children
with RD, control group; RD-TRT = children with RD, treatment group.
measures. Thus, a 2 (ability group) × 2 (condition) × 2 (test wave: pretest vs. posttest) MANOVA was computed on Listening Sentence Span item scores, with repeated measures on the last factor. As shown in Table 5 and Figure 1b, a significant main effect emerged for group, Wilks’s $\Lambda = .77$, $F(1, 54) = 7.90, p < .001, \eta^2 = .33$, and testing wave, Wilks’s $\Lambda = .80$, $F(1, 54) = 13.31, p < .001, \eta^2 = .20$. No other significant effects emerged, $ps > .05$. Item recall scores were lower for children with RD than for children without RD, and posttest scores were higher at posttest than at pretest. No significant effects emerged for treatment conditions, $ps > .05$. A similar analysis was computed on process scores. The only significant effect to emerge related to test wave, Wilks’s $\Lambda = .82$, $F(1, 54) = 11.79, p < .001, \eta^2 = .18$. The results indicated that process scores were higher for the posttest than for the pretest.

Although no significant treatment effects emerged related to item scores or process scores, we determined if the treatment conditions were influenced by the trade-off differences scores (posttest – pretest) for the Listening Sentence Span task. A significant effect emerged for condition, $F(1, 54) = 4.06, p < .05, \eta^2 = .08$. No other significant effects emerged, $ps > .05$. The results indicated that trade-off scores were higher for the control conditions ($M = 0.15, SD = 0.37$) than for the treatment conditions ($M = -0.02, SD = 0.30$). Thus, treatment conditions influenced (placed greater demands on) the trade-off (interchange) between process and storage for both ability groups.

**Correlations.** The next analysis determined whether the correlations between pretest and posttest were reduced in the rehearsal condition when compared to the control condition. The correlations between pretest and posttest for the computation span test with achievement and listening span as a function of treatment are shown in Table 6. Because of the sample size and range in the scores, Spearman rho (rank order) correlations were computed. Also included in the analysis was the biserial correlation for ability group. Although several of the correlations were significant, dependent $t$ tests between pretest and posttest scores within conditions yielded no significant effects related to the magnitude of the correlation, $ps > .05$. In addition, the magnitude of the biserial correlations between control and rehearsal conditions at posttest for the independent sample were statistically comparable, all $ps > .05$. Thus, we found no support for the assumption that variance between high (children without RD) and low (children with RD) WM groups was reduced as a function of strategy training.

**Hierarchical regression analyses.** Hierarchical regression analyses were conducted to determine if effects related to treatment on the transfer and achievement measures were merely a function of pretest performance. Because of the sample-to-task ratio, only four predictor variables were considered in the analysis (pretest, condition, posttest, and Posttest × Condition). Scores for the criterion measures used in these analyses were raw scores from the Reading and Math subtests of the WRAT-III and CST. Scores for the predictor variables were item recall. Of interest was whether posttest scores contributed significant variance to achievement or transfer when pretest scores were partialed from the analysis. The intercorrelations for the total sample for this analysis are shown in Table 7.

| Table 6. Spearman Rho Correlations Between Achievement and Transfer Measures With Operation Span (item recalls) Pretest Item, Posttest Item, and Ability Group |
|----------------------------------|---------------------------------|---------------------------------|---------------------------------|
|                                  | Control                         | Rehearsal                       |
|                                  | Pre | Post | Group | Pre | Post | Group |
| WRAT-III Reading                 | .64** | .48** | -.85** | .67** | .41 | -.84** |
| WRAT-III Math                   | .43* | .30 | -.49** | .47** | .28 | -.53** |
| STAR Reading                     | .28 | .26 | -.31 | .63** | .62** | -.56** |
| STAR Math                        | .34 | .27 | -.19 | .35 | .59** | -.26 |
| Transfer measure                |     |     |       |     |     |       |
| Listening span                  |     |     |       |     |     |       |
| Pretest                         | .29 | .23 | -.30 | .43* | .47** | -.18 |
| Posttest                        | .34 | .10 | -.45** | .53** | .36** | -.46* |

Note: WRAT-III = Wide Range Achievement Test; STAR = California Standardized Testing and Reporting program.  *p < .05, **p < .01.
Table 7. Intercorrelations Between Achievement, Item Recall for Working Memory Tasks, and Treatment for Experiment 2 (total sample)

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
</tr>
</thead>
<tbody>
<tr>
<td>Raven</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WRAT-III Reading</td>
<td>.14</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WRAT-III Math</td>
<td>.17</td>
<td>.64**</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Listening Sentence Span pretest</td>
<td>.06</td>
<td>.27*</td>
<td>.23</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Listening Sentence Span posttest</td>
<td>-.16</td>
<td>.47*</td>
<td>.32*</td>
<td>.36**</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Operation Span pretest</td>
<td>.15</td>
<td>.59**</td>
<td>.42**</td>
<td>.30*</td>
<td>.40**</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Operation Span posttest</td>
<td>.03</td>
<td>.25</td>
<td>.17</td>
<td>.28*</td>
<td>.35**</td>
<td>.42**</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Treatment condition</td>
<td>-.13</td>
<td>-.08</td>
<td>-.04</td>
<td>.03</td>
<td>.13</td>
<td>-.16</td>
<td>.63**</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Group a</td>
<td>-.07</td>
<td>-.80**</td>
<td>-.50**</td>
<td>-.22</td>
<td>-.47**</td>
<td>-.63**</td>
<td>-.36**</td>
<td>-.10</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CST Reading b</td>
<td>.12</td>
<td>.34**</td>
<td>.29*</td>
<td>.04</td>
<td>.13</td>
<td>.43</td>
<td>.12</td>
<td>-.24</td>
<td>-.29</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CST Math b</td>
<td>.20</td>
<td>.08</td>
<td>.21</td>
<td>.10</td>
<td>.12</td>
<td>.34</td>
<td>.34*</td>
<td>.02</td>
<td>-.12</td>
<td>.77**</td>
<td></td>
</tr>
</tbody>
</table>

M: 30.23 33.95 35.51 10.72 12.50 22.56 27.22 1.51 0.50 46.00 39.64
SD: 2.84 4.83 5.75 3.36 3.20 6.17 6.35 0.50 0.50 18.31 18.41

Note: Raven = Raven Colored Progressive Matrices raw scores; WRAT-III = Wide Range Achievement Test raw scores; group = children with reading disabilities vs. children without reading disabilities group; CST = California Standards Test.

aSample size = 58.
bSample size = 50.
*p < .05. **p < .01.

Table 8. Hierarchical Regression Predicting Reading and Math Scores on the Wide Range Achievement Test (WRAT-III) and the California Standardized Testing and Reporting (STAR) Test and Listening Sentence Span

<table>
<thead>
<tr>
<th></th>
<th>WRAT-III Reading</th>
<th></th>
<th>WRAT-III Math</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>R^2</td>
<td>ΔR^2</td>
<td>F Ratio</td>
<td>R^2</td>
</tr>
<tr>
<td>Model 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pretest</td>
<td>.35</td>
<td></td>
<td></td>
<td>.18</td>
</tr>
<tr>
<td>Model 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pretest</td>
<td>.36</td>
<td></td>
<td></td>
<td>.18</td>
</tr>
<tr>
<td>Condition</td>
<td>.36</td>
<td></td>
<td>.03</td>
<td>.18</td>
</tr>
<tr>
<td>Posttest</td>
<td>.36</td>
<td></td>
<td>.04</td>
<td>.18</td>
</tr>
<tr>
<td>Posttest × Condition</td>
<td>.37</td>
<td>.01</td>
<td>.31</td>
<td>.19</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>STAR Reading</th>
<th></th>
<th>STAR Math</th>
<th></th>
<th>Listening Sentence Span</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>R^2</td>
<td>ΔR^2</td>
<td>F Ratio</td>
<td>R^2</td>
<td>ΔR^2</td>
<td>F Ratio</td>
</tr>
<tr>
<td>Model 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pretest</td>
<td>.18</td>
<td></td>
<td></td>
<td>.12</td>
<td></td>
<td>.65**</td>
</tr>
<tr>
<td>Model 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pretest</td>
<td>.18</td>
<td></td>
<td></td>
<td>.12</td>
<td></td>
<td>.69**</td>
</tr>
<tr>
<td>Condition</td>
<td>.20</td>
<td>.02</td>
<td>.24</td>
<td>.14</td>
<td>.02</td>
<td>1.13</td>
</tr>
<tr>
<td>Posttest</td>
<td>.21</td>
<td>.01</td>
<td>1.18</td>
<td>.18</td>
<td>.04</td>
<td>.24</td>
</tr>
<tr>
<td>Posttest × Condition</td>
<td>.24</td>
<td>.03</td>
<td>2.22</td>
<td>.23</td>
<td>.05</td>
<td>2.61</td>
</tr>
</tbody>
</table>

*p < .05. **p < .01. ***p < .001.

Model 2 considered the predictability of posttest Operation Span scores as a function of treatment on achievement (reading and math) and the transfer measure (listening span). As indicated at the top of Table 8, the pretest, condition, posttest, and Condition × Posttest interaction variables accounted for 37%, F(4, 53) = 7.38, p < .0001, of the variance in reading and 19%, F(4, 53) = 3.04, p < .05, of the variance in math. As shown at the bottom of Table 8, pretest scores, condition, posttest scores, and Posttest × Condition scores accounted for 24%, F(4, 45) = 3.47, p < .0001, of the variance on the
The purpose of this study was to determine the role of strategy knowledge, processing efficiency, and capacity constraints on the WM performance in children with RD. Of particular interest was whether strategies reduce the variance in WM span between children with and without RD and/or whether processing efficiency or constraints drive the relationship between reading and WM. The hypothesis under investigation was whether children with RD (because of their low WM span) would benefit more from strategy instruction than do children without RD (children with high spans) because strategies allow them to compensate for processing constraints on WM. Taken together, the two experiments show that children with RD benefited significantly from strategy instruction (e.g., cuing and rehearsal), but such procedures failed to account for the covariation between WM and reading. The results of the two experiments will next be summarized addressing the major research questions that motivated the studies. This analysis is followed by considerations of alternative interpretations of the results as well as placing the results within the context of the current literature.

Does Strategy Knowledge Account for Reading Group Differences in WM?

There was weak support for the assumption that strategy knowledge underlies the relationship between WM and reading. More specifically, the results of Experiment 1 addressed the interplay between strategy knowledge, stability of strategy choices, and strategic procedures (cuing) designed to bolster WM performance. The central question was whether any of these aforementioned variables underlie the poor WM performance of children with RD. The results showed that regardless of the WM conditions, the WM performance of children with RD was inferior to that of children without RD. The results showed that although both verbal and visual-spatial WM performance in children with RD improved via cuing procedures, the results that best predicted reading performance were captured by constraints in WM performance. No unique variance was found related to predictions of reading performance as a function of individual differences in processing efficiency or the stability of strategy choices.

Does Strategy Training Reduce WM Differences Between Readers?

There was weak support of the assumption that strategy training reduces the variance in WM performance between reading groups. In contrast to Experiment 1, Experiment 2 provided direct training in strategy use. Whereas Experiment 1 used a scaffolding approach to cue children’s WM performance, Experiment 2 provided direct rehearsal training. As in Experiment 1, a major concern was whether the training effects were greater in children with RD than in children without RD. In contrast to Experiment 1, however, the study tested whether direct strategy training helped children with RD allocate WM resources more efficiently as well as reduce that variance in performance between the two reading groups.

In general, the results of Experiment 2 showed that rehearsal training significantly improved performance on the Operation Span measure for both reading groups. The regression analysis suggested that some transfer occurred as a function of treatment effects on the Listening Sentence Span transfer measure. However, the results related to the process and storage of the Operation Span task did not support the hypothesis that strategy training reduced the variance between reading groups.

Three findings support this conclusion. The first was that there was no significant Ability Group × Treatment interaction when an analysis was made of item scores, suggesting that the pattern of performance was comparable between the two ability groups. Although a significant interaction emerged for process and trade-off scores, this interaction showed that the differences in WM performance between reading groups increased. The second was that the magnitude of the correlations between pretest and WM and posttest and WM were statistically comparable. It was assumed that if strategy instruction reduces the variance between ability group, then two findings should have occurred: (a) The correlations between reading and WM
would be reduced at posttest compared to pretest, and (2) the interplay (trade-offs) between process and storage should be higher scores (showing reduction in demands) for children with RD than for children without RD. As shown in Table 6, the covariation between WM and reading was not reduced as a function of treatment. The results related to the second assumption are addressed in the third finding. The third finding was that the trade-off scores that reflected the interplay between process and storage were not in the expected direction. Although there was evidence that trade-offs between storage and processing were influenced by training conditions, the trade-offs were more negative for children with RD in the treatment condition, suggesting that training increased rather than reduced concurrent demands on storage and processing.

**What Accounts for the Variability in WM Performance?**

Given that individual differences in strategies knowledge or use do not fully account for WM span scores, the question emerges: What accounts for the variability in WM performance between reading groups? To answer this question, each experiment will be considered in detail as will alternative interpretations to the findings.

**Strategy, efficiency, and/or capacity?** Experiment 1 found that although ability group differences in favor of children without RD emerged across memory conditions, no differences were found between the groups in terms of strategy choices. Although the instability of strategy choices was more frequent in children with RD than in those without RD, the regression analysis found that only measures related to demands on processing capacity contributed significant independent variance to reading. Entering the capacity variable into the regression analysis eliminated the significant contribution of strategy stability in predicting reading comprehension. In addition, no support was found for the notion that reading group differences were related to processing efficiency. That is, the number of cues necessary to establish asymptotic performance, an assumed measure of processing efficiency, was not an important discriminator among the ability groups. In addition, children with RD were inferior to chronological matched children on conditions of high processing demands (maintenance) and conditions that enhance access to previously stored information (gain conditions).

In general, these results suggest that the important variance related to reading groups is related to processing constraints of a limited capacity system. Two findings support this interpretation. First, although the results show that cues substantially improve WM performance in children with RD (as shown by mean score improvement from initial to gain performance), children with RD remained at a clear disadvantage when compared to skilled readers in WM performance across gain and maintenance conditions. These low WM span scores emerged across both verbal and visual-spatial WM tasks, suggesting that the source of reading group differences reflects generalized processing constraints. Second, children without RD experience less reduction in performance on the maintenance condition than do children with RD. Thus, not only did children with RD have smaller WM spans than children without RD on initial and gain conditions, but they also experienced greater constraints on capacity (i.e., performance costs) under high demand (maintenance) conditions. Further, these constraints were statistically comparable across the verbal and visual-spatial tasks.

There are alternative interpretations of the results in Experiment 1. One considered is that strategy choices do not in fact reflect strategy knowledge but merely reflect an interference condition quite apart from actual declarative knowledge. This alternative interpretation suggests that children with RD are less resistant to interference (also see Chiappe, Hasher, & Siegel, 2000, for discussion of this model when applied to RD) than are the other ability groups. For example, it may be argued that the strategy questions and process questions for the verbal and visual-spatial WM tasks constitute a very temporary competing condition with storage. Further, the WM tasks vary in the number of processing questions and strategies to which the respondents are exposed, and the pervasiveness of RD children’s poor performance across such diverse measures may reflect a general interference condition. As a consequence, children with RD have difficulty preventing unnecessary information from entering WM and, therefore, are more likely to consider alternative interpretations of material (such as asked for in the strategy and processing questions) that are not central to the task. This interpretation fits within several recent models that explain individual differences in memory performance as related to inhibitory mechanisms (e.g., Hasher, Lustig, & Zacks, 2007), without positing some form of a strategy or capacity deficit.

Although we see the above model as a viable alternative to the results, we have three reservations. First, if strategy choice served only as an interference task, then such choices would be random and unrelated to task performance. That is, those students who chose particular strategies would not be expected to have better WM performance than those who chose strategies randomly. The improvements in WM scores in Experiment 1 associated with consistent strategy selection suggest this was not the case. Second, related to the processing questions, only respondents who answered the process question correctly were analyzed. That is, if a process question was missed, the child was not asked to recall previously stored information.
information. This procedure is different from those of previous studies (e.g., Daneman & Carpenter, 1980), which have allowed dissociation between the process question (i.e., it is not necessary for respondents to answer the process question correctly) and the retrieval question. This control provides feedback to respondents related to the interpretation and/or relevance of the material to be remembered plus the strategies selected. Thus, there was an experimenter-imposed association between the process and retrieval for the same set of inputs. Further, if children with RD suffer more interference (i.e., diminished inhibition in that a large number of traces are simultaneously active) than do children without RD, one would expect the probing to narrow significantly the alternative interpretations of items in memory when compared with chronological age matched children. That is, inefficiencies in inhibiting traces or competition effects should be reduced more in RD than in chronological age matched children. Further, one would predict that a procedure that gives feedback on the relevancy of a response should lead to a substantial increase in memory performance in the group with the diminished inhibitory efficiency. An analysis of effect sizes related to the gain condition, however, indicates statistically comparable changes between the ability groups.

**Span variability: strategy use, processing, and/or storage?**

Experiment 2 investigated whether implementation of procedural knowledge underlies the correlations between WM and reading. Previous studies have shown that training improved WM performance in college students (McNamara & Scott, 2001; Turley-Ames & Whitfield, 2003). The studies found that strategy intervention improved posttest WM performance and moderated the correlation between WM and reading comprehension. Experiment 2 determined whether individual differences in strategies used during performance of WM tasks were at the heart of the relationship between children’s WM capacity and reading. Although Engle, Cantor, and Carullo (1992) failed to support the strategic allocation hypothesis with adult samples, the hypothesis had not been tested prior to this study with elementary school children with high and low span scores.

As in the Turley-Ames and Whitfield (2003) study, Experiment 2 found that the use of rehearsal strategy instruction positively influenced posttest WM span scores on the WM task for which strategy instruction was given. The strategy instruction yielded higher posttest scores for operation span than in the control condition. Both groups in this study (children with and without RD) benefited from strategy instruction; however, the gains were not greater for children with low spans (children with RD) than for children with high spans (children without RD). This finding does complement the Turley-Ames and Whitfield study, where low spans (assuming a parallel can be drawn between low span and RD) showed more benefit from strategy instruction than did high spans.

Despite the positive effects of strategy use on children’s operation span performance, the results do not provide strong support for the assumption that the relationship between WM and RD is related to strategy variables. The main finding on this issue was that correlation coefficients were statistically comparable between pretest WM and reading and between posttest WM and reading. Further, standard deviations for WM at pretest and posttest were in the same range, suggesting that item recall (i.e., span variability) was not reduced. Previous studies with adult samples (e.g., McNamara & Scott, 2001; Turley-Ames & Whitfield, 2003) have assumed that if strategies account for the relationship between WM and reading accuracy, then the correlation between WM and reading should be significantly reduced (decreased) at posttest when participants are systematically exposed to the same strategy instructions. Thus, we expected that correlations between pretest WM and reading accuracy should be significantly larger in magnitude than between posttest WM and reading. This correlation pattern was not expected to emerge for children in the control condition. Our results show that instructed children improved their span scores relative to the control or uninstructed conditions, but the scores following instruction were as related to reading as were scores for the uninstructed condition. These findings are comparable to those of Engle et al. (1992) showing that strategies do not drive the covariation between WM and reading ability. However, it is important to note that although some controls were placed on study time during the posttest, we are uncertain if participants in the control condition may have modified their approach to the task despite no instructions to do so. Thus, we cannot conclude that strategies were unrelated to the covariation between WM and reading. No doubt, the study would have benefited from procedures that use step-by-step strategy reports to validate strategy use (e.g., Dunlosky & Kane, 2007).

There are alternative interpretations as well as limitations to Experiment 2. First, the amount of practice participants received in Experiment 2 using the described strategies (approximately 10–15 minutes) could have limited the benefits of the strategy use. With more practice, the strategy might have been more effective and might have been used in other tasks. Second, strategy instruction was given to only one WM task. Although rehearsal strategy has been shown to provide greater benefit to adult learners than do semantic or imagery strategies (Turley-Ames & Whitfield, 2003), strategy instruction on other WM span tasks might have increased the possibility that participants would employ the strategy,
Given that the poor WM performance for children with RD is attributed to capacity limitations, an obvious question emerges as to why rehearsal training helped this reading group. Our best explanation is that rehearsal training is easy to learn and requires little practice to master, which means it is not resource demanding. It is possible that all students had adequate resources to enact this strategy without placing demands on WM. The rehearsal strategy may have provided a technique that allowed students to focus on the relevant aspects of the task without being distracted by irrelevant information. Rehearsing the TBR words repeatedly might have activated the relevant information while preventing irrelevant information from interfering with the WM task. However, there does not appear to be convincing evidence that children with RD benefited more from strategy instruction, as exhibited by the differences between pretest and posttest Operation Span scores. Thus, low-span participants (children with RD) did not benefit more from rehearsal strategy instruction than did high-span participants. The reason for not acquiring such an interaction, however, may be because low-span children have been already using a strategy prior to strategy instruction. Thus, strategy training did not compensate for individual differences in WM capacity.

Some comment on the transfer effects is appropriate here. As indicated in the regression analysis for Experiment 2, there was an improvement in the Listening Sentence Span scores at posttest. One explanation is that the Listening Sentence Span task is a verbal measure, and the rehearsal procedure may be best applied to tasks of some verbal complexity. The Listening Sentence Span and the Operation Span are both verbal measures, but the operation task also draws on math skills, and math skills have been linked to verbal WM (Swanson & Jerman, 2006). Thus, although the results indicate that strategy instruction may transfer within domains (verbal WM tasks), the results showed no significant change in correlations between WM and reading or WM and math. There is positive evidence emerging, however, that WM performance can be influenced instructionally (e.g., Klingberg et al., 2005; Klingberg, Forssberg, & Westberg, 2002; Thorell, Lindqvist, Nutley, Bohlin, & Klingberg, 2009). Unfortunately, the evidence is not clear as to whether increases in WM have a direct influence on high-order skills such as reading comprehension. To test such transfer it will be necessary to show that the WM tasks share similar processes to reading in children with RD (e.g., constraints in capacity). This research has not been done so far. There is evidence suggesting, however, that transfer effects have emerged for children with ADHD. For example, Klingberg et al. (2002) found that WM training with ADHD children generalized to measures of fluid intelligence. Because WM forms the basis of some measures of fluid intelligence (e.g., Kyllonen & Christal, 1990), a general improvement in WM performance would be expected to lead to improvements on functions and tasks that rely on the same WM processes. Thus, the transfer effects we found on the Listening Sentence Span task appear related to both tasks sharing similar process.

In summary, the results of two experiments support the notion that children with RD have comparable strategy knowledge about WM tasks as do children without RD. Like children without RD, the stability of strategy choices is most directly related to span performance. Likewise, children with and without RD benefited from cued and rehearsal training conditions relative to control conditions. However, neither cued nor rehearsal training conditions allowed children with RD to improve their WM performance on a par with children without RD. In addition, measures of processing demands on WM capacity were more likely than processing efficiency and strategies to capture the source of reading group differences on WM tasks. However, it is important to note that although WM improved in both experiments, differences in WM between groups were not reduced. Both groups differed initially in WM in both experiments, suggesting that it is impossible to determine whether strategy differences were related to differences in WM or in RD. Thus, children with RD with different spans may respond differently to training conditions.

**Funding and Acknowledgments**

Janette Klingner served as the action editor for this manuscript. We thank three anonymous reviewers for their incisive comments on an earlier version. The research was partially supported by an Institute of Education Sciences grant (USDE R305H020055), U.S. Department of Education, and by Peloy Funds awarded to H. Lee Swanson. We are especially thankful to the Fontana School District and Riverside School District for data collection.

**Notes**

1. We assume that the maintenance condition reflects the storage capacity left after processing demands are accounted for. In support of this assumption, our previous work shows that the maintenance condition captures processes independent of those tapped in
the initial and gain conditions. For example, previous studies have shown that the maintenance condition predicts unique variance related to reading beyond that of span scores derived from the initial and gain condition (Swanson, Ashbaker, & Lee, 1996). Further, reading group differences emerge in the maintenance condition when the influence of initial span scores is partialed out (Swanson et al., 1996, Experiment 1). In addition, span scores from the maintenance condition predict age-related performance in children better than span scores from the initial and gain conditions (Swanson et al., 1996, Experiment 2). Several studies (e.g., see Unsworth & Engle, 2007, for a review) suggest that deficits in the ability to maintain information are a consequence of a limited capacity system. Therefore, failure to retrieve information in the maintenance condition can be partly attributed to demands placed on a limited capacity system. Of course, other influences may be operating, but it seems reasonable to assume that if reading group differences are related to WM capacity then RD readers will have fewer resources available to them to maintain and/or activate old (previously accessed) information than will skilled readers when processing demands are accounted for.

2. For the Digit-Sentence Span task, Swanson (1995) found in his standardization study (N = 968) that participants who selected a rehearsal strategy (in contrast to clustering, association, and elaboration) yielded significantly higher span scores than those participants selecting other strategies. Likewise, for a visual-spatial WM task (N = 957), the Mapping/Direction task, he found that participants selecting a sectional strategy (focus on recalling sections of a matrix) yielded higher span scores than those of participants who selected more global visual strategies.

3. Rehearsal strategies are more commonly associated with short-term memory (STM) tasks than with WM tasks. Clearly both WM and STM tasks involve sharing some common activities on the participant’s part. For example, both STM and WM tasks invoke controlled processes such as rehearsal (e.g., see Gathercole, 1998, for a review). However, controlled processing on WM tasks emerges in the context of high demands on attention (e.g., maintaining a memory trace in the face of interference) and the drawing of resources from the executive system (see Engle, Tuholski, Laughlin, & Conway, 1999, pp. 311-312, for discussion). Instructions in controlled processing emphasize maintaining information in the face of interference. Interference reflects competing memory traces that draw away from the targeted memory trace. In contrast, controlled processing on STM tasks attempts to maintain memory traces above some critical threshold (Cowan, 1995). This maintenance does not directly draw resources from the central executive system. Instructions in controlled processing may emphasize perceptual grouping or chunking skills, skills at phonological coding, and rehearsal speed (see Engle et al., 1999, for a review).

4. A complete discussion and example of probing procedures are provided in Swanson (1993b, 2003). To summarize, the “bow shaped curve,” commonly found in episodic memory studies, provided the basis for ordering a series of cues from implicit to explicit information. Cues are administered based on the type of error made (i.e., whether the error is related to recency, primacy, or middle items), and cuing procedures continue until all targeted items cannot be recalled. The order of cues was based on the assumption that the first cue provides information about the final items because these items are the least susceptible to interference. The second cue was assumed to provide information about the primacy (first) items because they are the most reliant on long-term memory processes. The third cue provides additional information about the middle-presented items because these items are the most susceptible to interference and storage limitations. Finally, if the participant fails to benefit from any of the previous three cues, all the items are repeated and retested. Probing procedures continue until all targeted items cannot be accessed (recalled). Because children were only probed about items for which they answered the process question correctly, it was assumed that poor item retrieval is attributable to item accessibility rather than to items not being adequately stored.

5. As indicated by one anonymous reviewer, we have no information as to whether the picture selection influenced the children’s approach to the WM task. We merely have measures of the consistency of choice. Previous work (Swanson, 1993b, pp. 109; Swanson et al., 1996, p. 258) has suggested that strategy scores do not differentiate between skilled and less skilled readers. Thus, although it cannot be discounted that strategy selection reflects an interference in processing, previous studies (Swanson, 1993b; Swanson et al., 1996) have shown that choices are not random (some strategies are selected more than others) and that the choices coincide with the task demands (rehearsal is selected for digit tasks, whereas clustering for semantic processing tasks).

6. As indicated by one reviewer, the outcomes of training may have differed if the training had included a similar constructed task that included reading. We have no data on this point.
Appendix

Operation Span Measure

Student Administration Instructions:

Say: “This is a memory task. You will be shown math problems, one at a time. You will read the problems aloud and give the answers. I will record your answers. After each math problem is answered, you will be shown a word to remember. You are to read the word aloud before you are shown the next math problem. When I say, ‘Recall words,’ you are to tell me all the words in order. You will be completing this task in sets. The first sets will have two problems, the next sets 3 problems, and so on up to sets with 5 problems. When I say, ‘Recall words,’ you will have as much time as you need to recall the words. You may guess if you are not sure. We will do some practice ones first.”

Version 1

Practice Set

1 + 1 = _____ to ______________________
7 + 3 = _____ me _____________________
0 + 5 = _____ by ______________________
7 – 2 = _____ and _____________________
6 + 1 = _____ about ___________________
0 + 2 = _____ was _____________________

Level 2

5 + 9 = _____ we _____________________
2 + 3 = _____ had _____________________
4 + 7 = _____ come ___________________
9 – 8 = _____ that _____________________

Level 3

0 + 4 = _____ if _______________________
3 + 5 = _____ get ______________________
9 + 3 = _____ in _______________________
2 + 4 = _____ some ____________________
8 + 1 = _____ or _______________________
1 + 2 = _____ not _______________________
References


Kyllonen, P. C., & Christal, R. E. (1990). Reasoning ability is (little) or more than working-memory capacity?! Intelligence, 14, 389–433.


About the Authors

H. Lee Swanson, PhD, is professor in educational psychology and special education at the University of California–Riverside. He received his doctoral degree from the University of New Mexico and conducted postdoctoral work at UCLA. His research interests are in cognitive processes, intelligence, and dynamic testing as it applies to children with learning disabilities.

Pam Kehler recently completed her PhD in special education at the University of California–Riverside. She currently is a resource room teacher and evaluator in the Riverside School District. Her research interests are in the area of learning disabilities, instruction, and assessment.

Olga Jerman recently completed her PhD in educational psychology at the University of California–Riverside. She currently serves as the research director at the Frostig School in Pasadena. Her research interests are in the area of working memory, problem solving, instruction, meta-analysis, and longitudinal research designs.