Superiority of variable to repeated practice in transfer on anagram solution

Michael K. Goode
Washington University, St. Louis, Missouri

Lisa Geraci
Texas A&M University, College Station, Texas

and

Henry L. Roediger III
Washington University, St. Louis, Missouri

Previous research in motor learning shows that practicing variations of a task (variable practice) leads to better transfer than does repeatedly practicing the exact same task (repeated practice). In contrast, research on priming using verbal materials shows that performance on a test improves to the extent that the material at learning and test overlap. We tested whether variability in practice conditions can lead to improved performance with the verbal priming task of anagram solution. Participants practiced solving anagrams, either repeatedly solving the same anagram that was later tested, repeatedly solving a different anagram from the one that was later tested, or solving different variations of the anagram that was later tested. We found that this last condition—variable practice on different versions of an anagram—led to improved test performance in relation to repeated practice, even when the test anagram was the one that had been repeatedly practiced. This finding aligns results from the motor learning literature with a higher level cognitive task: anagram solution. Shea and Kohl's (1991) hypothesis, arguing that varied practice may lead to greater elaborative processing than does repeated practice, provides one account of the results.

That variability of practice enhances learning is well established in the motor learning literature (e.g., Wulf & Schmidt, 1988). Practicing variations on a skill leads to better performance on transfer tasks. This is true even for perceptual learning tasks (Lively, Logan, & Pisoni, 1993). However, there has been little interest in whether this finding applies to learning and retention of verbal materials. In general, verbal learning and motor learning research have not always converged. However, as Schmidt and Bjork (1992) have argued, several phenomena occurring in both motor learning and verbal learning do converge on similar principles. The present experiment aims to examine a current gap between the two literatures in terms of the effect of variability of practice on later transfer.

Kerr and Booth (1978) examined the effect of variability of practice on motor skill learning and made a surprising finding. They investigated the effect of practicing throwing balls into buckets from variable distances. They found that children who practiced at variable distances (2 or 4 ft) outperformed those who practiced at one distance (3 ft), even though the test was from the same distance used for repeated practice (3 ft). That is, practice at variable distances produced greater transfer than practice at the exact same distance used in transfer. In a similar study, variable-plus-specific practice led to better performance on the specific task than did specific practice alone (Shea & Kohl, 1990). These studies showed that variable practice (or variable-plus-specific practice) can be superior to specific practice when the same specific retention test is used. Variable practice more often aids transfer to a novel task (e.g., Roller, Cohen, Kimball, & Bloomberg, 2001), particularly when practice order is randomized, rather than blocked, with similar items or skills being practiced together (Lee, Magill, & Weeks, 1985; Shea, Kohl, & Indermill, 1990).

The verbal learning literature has generally provided evidence for a contrary idea. The general finding is that transfer improves to the extent that there is increased overlap between the conditions at study and test. This is such a pervasive finding that it has been codified as the *encoding specificity principle* (Tulving & Thomson, 1973). A related principle, *transfer-appropriate processing*, states that performance at test improves to the extent that the processing engaged during a test overlaps with the processing engaged during study (Morris, Bransford, & Franks, 1977). This principle has been successfully applied to the literature on priming in implicit-memory tests (Blaxton, 1989; Roediger, Weldon, & Challis, 1989), in which memory is tested indirectly by examining improvement on a task given previous exposure to the components of the task (for a review, see Roediger & McDermott, 1993). Transfer on implicit tests is

H. L. Roediger III, roediger@wustl.edu

often highly specific and depends on recapitulating specific perceptual features, such as whether words were studied and tested in an identical font or modality (see Roediger & Srinivas, 1993; Schacter, Dobbins, & Schnyer, 2004, for reviews) or specific conceptual features (Geraci & Rajaram, 2004). In addition to traditional priming experiments, specificity of transfer has been shown in learning to read distorted (rotated or inverted) text (Kolers & Roediger, 1984).

The present study examined whether specificity or variability of practice would lead to better performance on a verbal test of anagram solution. We were interested in two questions relating to practice variability: First, does varied practice increase specific transfer to new anagrams from previously seen words—that is, does solving DOLOF, FOLOD, and OOFLD enhance participants' ability to solve LDOOF, in relation to three prior practice trials with LDOOF? (The answer is FLOOD.) Second, does varied practice lead to increased general transfer to solving new anagrams of new words—that is, does having solved DOLOF, FOLOD, and OOFLD lead to an increased ability to solve ABLEUVAL, in comparison with having solved LDOOF three times? We were also interested in whether test delay would interact with practice type. McAndrews and Moscovitch (1990) used a similar paradigm for other reasons, but they used only one practice session and found no difference between practice conditions.

In the present study, the participants were trained in three practice sessions before taking an immediate test, followed by a second test after a 48-h delay. There were three practice types: *same*, *varied*, and *different*. In the same-practice condition, the participants solved an anagram three times, and were then tested on that same anagram—that is, they tried to solve the anagram LDOOF three times, and were then tested on LDOOF. In the varied-practice condition, the participants solved three different anagrams of the same word and were tested on a fourth anagram of that word—that is, they solved DOLOF, FOLOD, and OOFLD, and were then tested on LDOOF. In the different-practice condition, the participants practiced solving a single anagram repeatedly, but they were tested on a different anagram of the word; for example, they tried to solve the anagram DOLOF three times, and were then tested on a different anagram of that word, LDOOF.

We used a randomized practice schedule because studies of motor learning generally have not found a benefit for variable practice over specific practice, unless practice is randomized, rather than blocked (Lee et al., 1985). Previous studies have shown that practicing tasks in a random order is also important in some cognitive learning tasks (Carlson & Yaure, 1990; Hiew, 1977; Lauer, Streby, & Battig, 1976).

METHOD

Participants

Forty-two participants from Washington University in St. Louis and 38 from Texas A&M University completed the experiment for partial course credit or pay. A total of 8 participants (6 from Washington University and 2 from Texas A&M) were excluded from the analyses for not following directions to attempt to solve anagrams for the full 30 sec allotted for each anagram. Thus, 36 participants from each location were used in the analyses.

Norming and Materials

A norming study was conducted in order to create a set of appropriately difficult anagrams to demonstrate transfer. To create the set of anagrams, we used a pool of 120 words ranging from 5 to 11 letters in length. For each word, four single-solution anagrams were constructed. The participants ($n = 40$) were given one of the four anagrams for each of the 120 words to solve. Each anagram was completed by 10 participants, who were given 30 sec to solve each anagram. Words with average completion rates between .14 and .89 across all four anagrams (and with average standard deviations lower than .36) were used for the experiment. This left a total of 100 words (400 anagrams) to be used for the experiment. The remaining 20 words (80 anagrams) were used as fillers during the practice sessions of the experiment. No more than one filler anagram derived from each word was used with the same participant during the experiment.

The 100 critical words (400 critical anagrams) for the experiment were divided into four study lists of 25 words (for counterbalancing purposes, these groups will be described as A, B, C, and D). These groups of words were matched on average solution frequency ($M = 0.62$, $SD = 0.20$), as well as on average length ($M = 7.1$, $SD = 1.5$).

Design and Counterbalancing

The experiment used a $3 \times 2 \times 2$ mixed design, with practice condition (*same*, *different*, and *varied*) serving as the between-participants variable, and test delay (immediate vs. delayed) and study status (studied vs. unstudied) serving as the within-participants variables. At each test, half of the tested words had been studied, and the other half had not. In the same-practice condition, the participants practiced solving a single anagram three times. In the different-practice condition, the participants also practiced solving the same anagram of a single word, but were tested on a different anagram of the word. In the varied-practice condition, the participants practiced three different anagrams of the same word and were tested on yet a fourth anagram of that word.

Half of the practiced anagrams were tested immediately, and the other half were tested after a 48-h delay. Thus, if the participants had studied the A-list and B-list anagrams (as well as the filler anagrams), then the immediate test would include studied A-list anagrams and unstudied C-list anagrams, whereas the delayed test would include studied B-list anagrams and unstudied D-list anagrams. Lists were counterbalanced across study status and test delay according to a balanced Latin square design, so that each list served in each condition once (practiced, tested immediately; practiced, tested at a 48-h delay; nonpracticed, tested immediately; nonpracticed, tested at a 48-h delay). This counterbalancing scheme produced four versions of the experiment for each of the three practice conditions, leading to a total of 12 between-participants conditions.

Procedure

The participants were tested either individually or in groups of up to four people. After giving informed consent, they began the experiment, which was conducted on computers. Instructions were displayed on the computer screen while the experimenter read them aloud. The participants were told to rearrange each anagram into an English word. They were given two practice anagrams to solve before they began the experimental anagrams and were given a maximum of 30 sec to input the solution to each anagram by typing their responses. If they solved the anagram in less than 30 sec, they could proceed to the next anagram. The participants were not given the correct answer if they were unable to solve an anagram in the allotted time. This was done to minimize the chance that performance would be driven by recall of previous solutions and not by improvements in the skill of solving the anagrams. Although specific feedback was not given, the participants knew when they could not complete an anagram, so *knowledge of results* (as it is called in motor learning research) was given.

The study session was divided into three blocks, with short breaks in between. During each study block, 50 experimental anagrams and 21 filler anagrams were presented. Anagrams were randomly chosen for each participant. For example, in the different-practice condition,

one of four possible anagrams for the word AWFUL was presented (e.g., WALUF). That anagram was then repeated during the second and third study blocks, and a different anagram was presented during the test block (e.g., LUFAW). Separate groups of participants thus received different anagrams (in this case, either WALUF, LUFAW, FULAW, or ALFUW) of the same words in the *same* conditions. Within each study block, the order of presentation of the anagrams was randomized.

There was a 3-min break between study blocks, during which the participants played a video game. After the three study blocks, there was another brief break, followed by the immediate test. On this test, the participants were given 25 studied anagrams and 25 unstudied anagrams, for a total of 50 anagrams. After finishing the test, the participants were instructed to return 2 days later. During the second session, they completed 25 anagrams repeated from the three previous practice sessions, as well as 25 new anagrams. We measured both accuracy and latency of correct anagram solutions. Response latency was measured from the time the anagram was first presented until the time that the participant solved the anagram (and pressed "Enter" to continue) or when the 30-sec time limit elapsed. The participants were thanked for their participation and debriefed.

RESULTS

For the following analyses, all results were significant at the .05 level. In addition to standard inferential statistics, p_{rep} was calculated for all statistical tests (Killeen, 2005). The statistical tests we conducted used difference scores (the difference in performance on studied and unstudied anagrams), but the results did not change when we compared only the test performance across conditions. Since the participants were tested at two locations, we analyzed performance during practice and test as a function of location. There was no main effect of location on the proportion of anagrams solved, nor were there any interactions between location and practice session, test session, or practice condition. Therefore, all analyses collapsed across location.

Performance During Practice Sessions

We examined each participant's learning across the three practice sessions and the mean proportion of studied anagrams solved, as a function of practice session. The results are shown in Figure 1, revealing a strong practice effect that was greater for the varied- than for the repeated-practice conditions. A 3 × 3 within-participants ANOVA revealed a significant main effect of practice [$F(2,138) = 225.98$, $MS_e = 0.98$, $p_{rep} > .99$, $\eta_p^2 = .77$], but no main effect of practice condition [$F(2,69) = 1.30$, $MS_e = 0.05$, $p = .28$, $p_{rep} = .78$, $\eta_p^2 = .04$]. However, the interaction between practice session and practice condition was significant [$F(4,138) = 3.85$, $MS_e = 0.01$, $p_{rep} = .98$, $\eta_p^2 = .08$], because participants who had received *varied* practice were more accurate than those who had received *same* or *different* practice by the third practice session.

We also examined changes in response latencies to solve each anagram across practice sessions, as shown in Figure 2. The participants became increasingly fast in all conditions, but especially in the *same* and *different* conditions, in which the anagram stayed the same across the three practice sessions. A 3 × 3 within-participants ANOVA revealed a significant decrease in response latency across practice sessions [$F(2,138) = 252.01$, $MS_e = 354.3$, $p_{rep} > .99$, $\eta_p^2 = .79$], but no main effect of practice condition ($F < 1$). There was a significant interaction between practice condition and
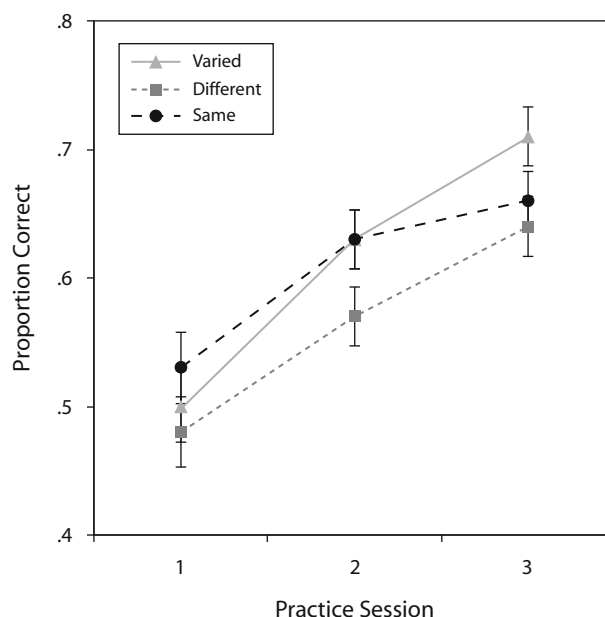

Figure 1. Proportion of anagrams correctly solved, as a function of practice condition and practice session.

practice session [$F(4,138) = 3.67$, $MS_e = 5.17$, $p_{rep} = .96$, $\eta_p^2 = .10$], indicating, as noted, that the participants in the repeated-practice conditions became faster across sessions than did those in the varied-practice condition.

Transfer

Table 1 provides the proportion of anagrams solved for each condition at the two test delays for both studied and unstudied items, along with transfer (calculated as studied proportion correct minus unstudied proportion correct at test). On the immediate test, varied practice led to greater transfer to solving a new anagram than did repeated prac-
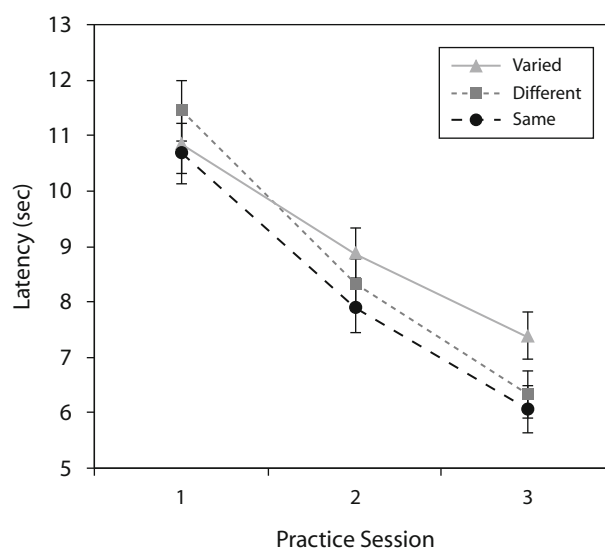

Figure 2. Latency to correctly solve anagrams, as a function of practice condition and practice session.

Table 1
Proportion of Anagrams Solved, As a Function of
Test Delay, Practice Condition, and Study Status

Study Status	Immediate Test						Delayed Test					
	Same		Different		Varied		Same		Different		Varied	
	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD
Studied	.68	.13	.69	.11	.77	.12	.71	.13	.67	.12	.74	.12
Unstudied	.56	.17	.55	.14	.54	.12	.62	.16	.57	.14	.56	.16
Transfer	.12	.11	.14	.11	.23	.08	.09	.11	.10	.14	.18	.15

tice of the target anagram or repeated practice of one different anagram. A 3×2 ANOVA revealed a significant main effect of test delay, with participants showing more transfer on the immediate test than on the delayed test [$F(1,69) = 5.64$, $MS_e = 0.068$, $p_{rep} = .95$, $\eta_p^2 = .08$]. It also revealed a significant main effect of practice condition [$F(2,69) = 7.62$, $MS_e = 0.13$, $p_{rep} > .99$, $\eta_p^2 = .18$]. There was no interaction between test delay and practice condition. To examine the effect of practice condition, we conducted separate Scheffé post hoc tests for the immediate and delayed tests. For the immediate test, there was no difference between the same- and different-practice conditions, but participants in the varied-practice condition showed significantly more transfer than those in the same-practice condition ($p_{rep} = .98$) and in the different-practice condition ($p_{rep} = .95$). On the 48-h-delayed test, there were no significant differences in performance among the different practice conditions, although numerically, the greatest transfer again occurred in the varied-practice condition.

We also examined the effect of study and test condition on latency to complete anagrams. We calculated latency transfer scores by subtracting each participant's reaction time to solve the studied anagrams at test from his/her reaction time to solve the unstudied anagrams at test. Faster transfer reaction times indicate greater transfer. Table 2 shows the latencies as a function of practice condition. Greater transfer occurred on the immediate than on the delayed test. A 3×2 (practice condition $\times$ test delay) repeated measures ANOVA confirmed this by showing a significant effect of test delay [$F(1,69) = 50.85$, $MS_e = 1.50 \times 10^8$, $p_{rep} > .99$, $\eta_p^2 = .42$]. There was no significant effect of practice condition and no interaction between practice condition and test delay.

General Transfer

We examined the effect of solving particular anagrams on general transfer to anagrams of new words by analyzing the effect of practice condition on performance for unstudied anagrams, both on the immediate and delayed test. First, we analyzed anagram solution accuracy, as shown in

Table 1. The proportion of anagrams correctly completed for unstudied anagrams on both the immediate and delayed tests were compared to the proportion of anagrams solved during the first study session using separate paired-samples t tests. There was significant transfer of learning to the unstudied anagrams, with participants solving more unstudied test anagrams than first-session studied anagrams on both the immediate [$t(71) = 2.74$, $p_{rep} = .97$] and delayed [$t(71) = 5.81$, $p_{rep} > .99$] tests. To ascertain whether general transfer was influenced by practice condition, we conducted separate one-way ANOVAs on the data from the immediate and delayed tests. There were no significant differences in the proportion of anagrams solved, as a function of practice condition, on either the immediate [$F(2,69) = 0.08$, $p > .90$, $p_{rep} < .53$] or delayed [$F(2,69) = 1.06$, $p > .35$, $p_{rep} < .75$] test. Although varied practice improved specific transfer more than did the other practice conditions, general transfer was equivalent across practice conditions. Therefore, the benefit of varied practice on solving specific anagrams is not due to a nonspecific practice effect.

To further examine general transfer, we analyzed response latencies to correct, unstudied anagrams (see Table 2). Results showed significant transfer to unstudied anagrams; the participants solved unstudied test anagrams more quickly than they did first-session, studied anagrams, on both the immediate [$t(71) = 3.61$, $p_{rep} = .99$] and the delayed [$t(71) = 3.41$, $p_{rep} = .99$] test. We also conducted separate one-way ANOVAs on the response latency data from the immediate and delayed tests. Response latency was influenced by practice condition on the immediate test [$F(2,69) = 7.35$, $MS_e = 2.85 \times 10^7$, $p_{rep} = .99$]. Post hoc Scheffé tests revealed that participants in the same-practice condition were significantly faster at solving the anagrams than were participants in either the different-practice condition ($p_{rep} = .98$) or the varied-practice condition ($p_{rep} = .95$), but mean reaction times did not differ between the latter two groups. A one-way ANOVA revealed no significant effects of practice condition on latency for the delayed test [$F(2,69) = 2.03$, $p = .14$, $p_{rep} = .85$].

Table 2
Latency to Solve Anagrams (in Seconds),
As a Function of Test Delay, Practice Condition, and Study Status

Study Status	Immediate Test						Delayed Test					
	Same		Different		Varied		Same		Different		Varied	
	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD
Studied	4.59	1.89	6.66	1.96	6.22	2.05	7.35	1.56	8.22	1.42	8.30	2.31
Unstudied	9.38	1.96	10.44	2.25	9.85	3.20	9.63	2.78	10.48	1.96	9.74	2.71
Transfer	4.79	2.15	3.78	2.55	3.63	2.42	2.28	1.74	2.26	1.48	1.44	1.38

DISCUSSION

Variable practice led to improved transfer to new anagrams of practiced words in relation to repeated practice on one of those anagrams, even when the practice occurred with the target anagram used in the transfer phase. That is, after just three practice sessions, transfer to new anagrams was greater with variable practice than with repeated practice on the very anagrams used in the transfer test. This outcome conceptually replicates the Kerr and Booth (1978) study and, to our knowledge, is the first such demonstration using verbal materials.

The superiority of the variable-practice condition over the different-practice condition can be interpreted within several frameworks. Schmidt's (1975) schema theory, used to explain the benefit of variable practice in motor learning, can be adapted to verbal materials. Just as variable practice can help in learning a general motor schema, leading to better transfer to novel tasks, variable practice could lead to learning a lexical schema that can be used for solving anagrams. However, explaining how variable practice leads to greater performance than does repeated practice within this framework is more difficult. One idea, following on work by Battig (1979) and Shea and Kohl (1991), is that variable practice leads to a greater opportunity for distinctive, elaborative processing in relation to repeated practice. Battig proposed that increasing contextual interference during the learning of a cognitive task leads to improved retention. Successful performance on learning trials in the same- and different-practice conditions probably involved simply remembering the correct solution on many trials. The decreased response latencies over practice in the repeated-practice condition in relation to that in the variable-practice condition support this inference. Because the processing of these repeated anagrams was faster than the processing of new anagrams, there was less need for different (distinctive) processing of anagrams in these conditions than in the variable-practice condition. Variable practice therefore probably encouraged more distinctive or elaborative processing than did repeated practice, and hence led to greater transfer.

The present results appear to be at odds with much of the verbal memory literature, in which typical findings show that similarity or overlap of processing between practice and testing leads to enhanced performance. Our results demonstrate that variability of practice can improve transfer, at least on the task of completing anagrams.

At present, we cannot reconcile these results with the typical findings in the verbal memory literature.

REFERENCES

Battig, W. F. (1979). The flexibility of human memory. In L. S. Cermak & F. I. M. Craik (Eds.), *Levels of processing in human memory* (pp. 23-44). Hillsdale, NJ: Erlbaum.

Blaxton, T. (1989). Investigating dissociations among memory measures: Support for a transfer-appropriate processing framework. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **15**, 657-668.

Carlson, R. A., & Yaure, R. G. (1990). Practice schedules and the use of component skills in problem solving. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **16**, 484-496.

Geraci, L., & Rajaram, S. (2004). The distinctiveness effect in the absence of conscious recollection: Evidence from conceptual priming. *Journal of Memory & Language*, **51**, 217-230.

Hiew, C. C. (1977). Sequence effects in rule learning and conceptual generation. *American Journal of Psychology*, **90**, 207-218.

Kerr, R., & Booth, B. (1978). Specific and varied practice of motor skill. *Perceptual & Motor Skills*, **46**, 395-401.

Killeen, P. R. (2005). An alternative to null-hypothesis significance tests. *Psychological Science*, **16**, 345-353.

Kolers, P. A., & Roediger, H. L., III (1984). Procedures of mind. *Journal of Verbal Learning & Verbal Behavior*, **23**, 425-449.

Lauer, P. A., Streby, W. J., & Battig, W. F. (1976). The effects of alphabetic organization on the acquisition and delayed retention of semantically similar words. *Journal of Experimental Psychology: Human Learning & Memory*, **2**, 182-189.

Lee, T. D., Magill, R. A., & Weeks, D. J. (1985). Influence of practice schedule on testing schema theory predictions in adults. *Journal of Motor Behavior*, **17**, 283-299.

Lively, S. E., Logan, J. S., & Pisoni, D. B. (1993). Training Japanese listeners to identify English /r/ and /l/: II. The role of phonetic environment and talker variability in learning new perceptual categories. *Journal of the Acoustical Society of America*, **94**, 1242-1255.

McAndrews, M. P., & Moscovitch, M. (1990). Transfer effects in implicit tests of memory. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **16**, 772-788.

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

Roediger, H. L., III, & McDermott, K. B. (1993). Implicit memory in normal human subjects. In F. Boller & J. Grafman (Series Eds.) and H. Spinnler & F. Boller (Section Eds.), *Handbook of neuropsychology: Vol. 8. Section 11: Memory, dementia, perception of time, music, and faces* (pp. 63-131). Amsterdam: Elsevier.

Roediger, H. L., III, & Srinivas, K. (1993). Specificity of operations in perceptual priming. In P. Graf & M. E. J. Masson (Eds.), *Implicit memory: New directions in cognition, development, and neuropsychology* (pp. 17-48). Hillsdale, NJ: Erlbaum.

Roediger, H. L., III, Weldon, M. S., & Challis, B. H. (1989). Explaining dissociations between implicit and explicit measures of retention: A processing account. In H. L. Roediger III & F. I. M. Craik (Eds.), *Varieties of memory and consciousness: Essays in honour of Endel Tulving* (pp. 3-41). Hillsdale, NJ: Erlbaum.

Roller, C. A., Cohen, H. S., Kimball, K. T., & Bloomberg, J. J. (2001). Variable practice with lenses improves visuo-motor plasticity. *Cognitive Brain Research*, **12**, 341-352.

Schacter, D. L., Dobbins, I. G., & Schnyer, D. M. (2004). Specificity of priming: A cognitive neuroscience perspective. *Nature Reviews Neuroscience*, **5**, 853-862.

Schmidt, R. A. (1975). A schema theory of discrete motor skill learning. *Psychological Review*, **82**, 225-260.

Schmidt, R. A., & Bjork, R. A. (1992). New conceptualizations of practice: Common principles in three paradigms suggest new concepts for training. *Psychological Science*, **3**, 207-217.

Shea, C. H., & Kohl, R. (1990). Specificity and variability of practice. *Research Quarterly for Exercise & Sport*, **61**, 169-177.

Shea, C. H., & Kohl, R. (1991). Composition of practice: Influence on the retention of motor skills. *Research Quarterly for Exercise & Sport*, **62**, 187-195.

Shea, C. H., Kohl, R., & Indermill, C. (1990). Contextual interference: Contributions of practice. *Acta Psychologica*, **73**, 145-157.

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

Wulf, G., & Schmidt, R. A. (1988). Variability in practice: Facilitation in retention and transfer through schema formation or context effects? *Journal of Motor Behavior*, **20**, 133-149.