Journal of Experimental Psychology: Applied
2002, Vol. 8, No. 4, 222–232


1076-898X/02/$5.00 DOI: 10.1037/1076-898X.8.4.222

# Models of Performance in Learning Multisegment Movement Tasks: Consequences for Acquisition, Retention, and Judgments of Learning

Dominic A. Simon
McMaster University

Robert A. Bjork
University of California, Los Angeles

Participants learned different keystroke patterns, each requiring that a key sequence be struck in a prescribed time. Trials of a given pattern were either blocked or interleaved randomly with trials on the other patterns and before each trial modeled timing information was presented that either matched or mismatched the movement to be executed next. In acquisition, blocked practice and matching models supported better performance than did random practice and mismatching models. In retention, however, random practice and mismatching models were associated with superior learning. Judgments of learning made during practice were more in line with acquisition than with retention performance, providing further evidence that a learner's current ease of access to a motor skill is a poor indicator of learning benefit.

In the last several decades, a number of manipulations of practice and study conditions has been shown to have effects on acquisition performance that differ from the effects of those manipulations on learning—as assessed by delayed tests of retention and transfer (see reviews by Chamberlin & Lee, 1993; Christina & Bjork, 1991; Ghodsian, Bjork, & Benjamin, 1997; Schmidt & Bjork, 1992; Schmidt & Lee, 1999). Practice conditions that support higher levels of immediate performance do not always foster superior learning, and conditions that pose difficulties for the learner can enhance long-term retention and transfer (e.g., Schmidt, Young, Swinnen, & Shapiro, 1989; Shea & Morgan, 1979). Schmidt and Bjork (1992) reviewed a variety of evidence indicating that such phenomena are to be found in both the verbal and motor learning domains.

In the present study, we introduced a novel-practice manipulation: providing a matching or mismatching model of perfect performance before each trial on a given to-be-learned response sequence. In the matching case, the model illustrated perfect performances on the about-to-be-practiced response sequence, whereas in the mismatching case the model illustrated perfect performance on one of the other to-be-learned response sequences. For reasons we outline below, we anticipated that the mismatching-model condition would disrupt immediate performance but promote superior levels of learning, whereas the converse would be the case for the matching condition.

A central goal of the present research, however, beyond simply assessing the short-term and long-term consequences of providing such matching and mismatching models, was to investigate the attendant impressions that the participants developed about their own learning. Would learners' assessments of their learning be driven by their immediate performance or would they have insight into the benefits of the differing practice conditions as indicated by a delayed-performance test? Would the matching and mismatching models be seen as artificially enhancing or degrading performance, respectively, and thus be taken into account when predicting performance on a delayed test?

## Models of Performance as an Aid and Possible Hindrance to Learning

Providing a human model for learners is a common way to enhance their understanding and appreciation of the task to be learned (see McCullagh, 1993, for a review). Generally such a model, whether live or shown on a video or computer display, either demonstrates proper performance of the task or is a learner himself or herself, that is, someone who is also attempting to learn the desired movement pattern or skill in question. Modeled information can also be provided in various nonhuman forms. A simple example is the metronome used by musicians to provide them with the appropriate meter for a piece of music. Despite the established benefits of using them for learning, research by Lee, Wishart, Cunningham, and Carnahan (1997) suggests that the effects of models are not always positive. Indeed, they can act to nullify the effectiveness of other manipulations known to enhance long-term retention and transfer, more specifically, the benefits of random–interleaved conditions of practice, as in the so-called contextual-interference effect first suggested by Battig (1979, p. 24).

### *Random Versus Blocked Practice: Findings and Theories*

Broadly, the *contextual interference effect,* based on ideas from verbal learning (Battig, 1972, 1979), refers to the fact that arranging to-be-learned material in ways that create interference during learning can sometimes—unintuitively—enhance posttraining re-

Dominic A. Simon, Department of Kinesiology, McMaster University, Hamilton, Ontario, Canada; Robert A. Bjork, Department of Psychology, University of California, Los Angeles.

We thank Derek McAuley for helping to gather the data and Aaron Benjamin for comments on an earlier version of this article.

Correspondence concerning this article should be addressed to Dominic A. Simon, who is now at the Department of Psychology, Mail Stop Code 3452, New Mexico State University, P.O. Box 30001, Las Cruces, New Mexico 88003-8001. E-mail: domsimon@nmsu.edu

tention and transfer. By presenting a to-be-studied item in a context consisting of other to-be-learned materials, more processing interference occurs than if the item is presented surrounded, for example, by other presentations of that particular item. Following on the first demonstration of contextual interference in motor skills by Shea and Morgan (1979), studies of blocked versus random practice have typically involved a comparison of two training scenarios, each of which requires that multiple responses are to be learned. In blocked practice, all trials of a particular to-be-learned response are completed before moving on to begin practice on another response. In random practice, the to-be-learned responses are interleaved in a semirandom fashion. That is, the ordering of trials is not completely random, but each of the to-be-learned items appears a set number of times in random order within a subset of so many trials. Usually, there is also a prohibition against having more than two consecutive trials of the same response.

The standard finding (see, Chamberlin & Lee, 1993; Magill & Hall, 1990; Schmidt & Lee, 1999, for reviews) is that during practice itself, performance is superior under blocked- rather than random-practice conditions. That is, faster or more accurate performance of a multisegment movement task, for example, is seen during practice trials under a blocked than under a random-practice schedule (Lee & Magill, 1983; Lee et al., 1997; Shea & Morgan, 1979). In contrast, on delayed tests of retention and transfer, random-practice participants generally demonstrate superior performance to that of their blocked-practice counterparts. That is, faster or more accurate performance is now seen for those who practiced using a random rather than a blocked schedule (Lee & Magill, 1983; Lee et al., 1997; Shea & Morgan, 1979). Tasks in which this phenomenon has been observed include both motor and nonmotor activities as diverse as learning different badminton serves (Goode & Magill, 1986), shooting from different positions (Boyce & Del Rey, 1990), baseball batting practice (Hall, Domingues, & Cavazos, 1994), and learning electronic logic gate configurations and Boolean logic operations (Carlson, Sullivan, & Schneider, 1989; Carlson & Yaure, 1990).

Explanations as to why random–interleaved practice results in better long-term retention and transfer include Lee and Magill's (1983, 1985) reconstruction hypothesis and the elaboration and distinctiveness hypothesis proposed by Shea and Zimny (1983, 1988). The reconstruction hypothesis centers on the idea that in blocked practice, the learner need only generate the response pattern on the first trial of practice for a given pattern. Thereafter, feedback-dependent adjustments are made from trial to trial, but there is no need to reconstruct the response pattern from scratch. In random practice, however, trials of other response patterns must be generated between attempts at a given response. Owing to that requirement, learners in the random–interleaved practice regimen find it harder to maintain high levels of immediate performance, but their learning is enhanced via their greater experience at generating the appropriate response patterns anew.

According to the elaboration–distinctiveness hypothesis, random practice is thought to afford superior learning because it provides many more opportunities to compare and contrast the respective to-be-learned responses. As a consequence, more elaborated and more distinctive representations of the to-be-learned response patterns are thus thought to be formed, facilitating the learner's ability to produce the patterns at a retention delay.

### *Models Can Nullify the Benefits of Random Practice: Lee et al. (1997)*

In an experiment to test the reconstruction hypothesis of contextual interference, Lee et al. (1997) had participants learn to make three different keypress sequences, each with a unique movement-time (MT) goal. Apart from a standard blocked- and random-practice condition, they also included a novel condition in which random-practice trials were immediately preceded by presentations of a computer-generated model of the appropriate timing of the upcoming movement. The effect of the models was to make immediate performance similar to that of the blocked condition, despite the random order of actual practice. In delayed retention, also, the random-plus-model condition demonstrated performance similar to, if not slightly worse than, that of the standard blocked condition, and less accurate than that of the standard random condition. Thus, Lee et al.'s findings suggest that a matching model can override the difficulties intrinsic to random–interleaved practice conditions, but at the cost of also nullifying the learning benefits of such a schedule.

## Goals of the Present Research

### *Matching Versus Mismatching Models*

In the present experiment, we provided for some learners similar models to those used by Lee et al. (1997). Because these models were the same as the pattern that was to be executed immediately after their presentation, we refer to them as *matching models*. In another condition, mismatching models were used; that is, a model of perfect performance was also presented before each practice trial, but demonstrating perfect performance on a pattern other than the one about to be performed.

Participants learned three key sequences similar to those of Lee et al. (1997) in either a blocked or a random schedule. Given Lee et al.'s (1997) findings, we expected that matching models would support superior performance (smaller timing errors) than would mismatching models but would yield relatively poor learning, as measured by a delayed-retention test. In contrast, we expected that mismatching models would render immediate performance less accurate but would support superior learning. Practice schedules were expected to influence performance in the standard way: blocked practice superior during acquisition, but inferior in retention, when compared with random practice.

### *Subjective Assessments of Learning*

In addition to the performance data gathered in this experiment, we were also interested in determining the degree to which the schedule of practice, blocked or random, and the modeling conditions, matching or mismatching, would influence learners' metacognitive appraisals of their ongoing learning. In the domain of verbal learning and memory, there have been extensive investigations of the extent to which participants' predictions of their own future performance are or are not accurate (Metcalfe & Shimamura, 1994; for a review, see Bjork, 1999). Under some conditions, such judgments of learning (JOLs) are unreliable, and some researchers have even found negative correlations between such judgments and actual retention performance (e.g., Benjamin, Bjork, & Schwartz, 1998).

Self-assessments of one's proficiency in a given motor skill are necessary and prevalent in many real-world contexts—and such assessments presumably play a critical role not only in determining one's confidence that one would be able to perform the skill when needed (in a competition, in a crisis, on a test, etc.), but also in behavioral decisions, such as whether to seek further practice or instruction, whether to volunteer for certain assignments, and so forth—but research on metacognitive judgments in the motor domain has been relatively rare thus far (Simon & Bjork, 2001). It is particularly interesting to examine such judgments in the context of the present experiment because of the anticipated dissociation between acquisition performance and delayed-retention performance. If learners have insight into the degree to which their practice schedule and model combination are actually affecting their learning of the experimental task, then they should predict their retention performance to be superior for random practice than for blocked practice. Similarly, they should predict better retention performance for mismatching models than for matching models.

In contrast, if participants (falsely) assume that their current performance is a reliable indicator of learning, then participants who have the "benefit" of blocked practice and matching models during training are likely to predict superior retention performance, compared with participants who have the "burden" of random practice and mismatching models.

Our earlier findings (Simon & Bjork, 2001) suggested that the participants engaging in blocked practice would probably predict smaller retention errors than would actually occur, but, in the case of matching models, it seemed likely that the participants would be sensitive to the helpful role of such models in supporting immediate performance and thus adjust their predictions of error accordingly. Given that the participants were made aware that on the final criterion test such support would not be present, we suspected that participants in the matching-model condition would be likely to discount their current performance in predicting their future (criterion) performance.

If an instructor or coach were to show a learner the correct action before every attempt at a new movement pattern, thereby providing a nearly perfect, matching-performance model, it might be expected that the learner would anticipate deterioration in performance when the instructor model was no longer available. In contrast, learners exposed to mismatching models might be expected to feel that those models were interfering with their otherwise well-developed ability to generate the appropriate movement pattern. As such, they might be expected to yield predictions of better retention performance than those seen in the matching condition.

In general, the goal of the present research was to obtain performance data and the metacognitive prediction data that would speak to important applied questions of how to structure practice, and if necessary, the need to educate learners, and trainers, that their subjective impressions gained during practice may not accurately reflect their state of learning as demonstrated in a delayed test.

## Method

### *Participants*

The participants were 96 undergraduate students (56 women, 40 men: mean age = 19.8 yrs, *SD* = 3.4 yrs; range = 18–45 yrs) from the University of California, Los Angeles, who took part in exchange for credit in a psychology course. The participants were randomly assigned to one of four equal-sized groups for the acquisition phase of the experiment: blocked practice–matching models, blocked practice–mismatching models, random practice–matching models, and random practice–mismatching models. The participants were blind as to the hypotheses under test in the experiment.

### *Task and Apparatus*

The main experimental procedures were administered using a desktop computer. A customized software package was used to present the experimental trials, to collect participants' responses, and to record data for later analysis. The basic task involved learning to press three specific five-key sequences on the number pad of the computer keyboard (see Figure 1). The sequences were to be executed as close as possible to predetermined goal MTs. This task has been used previously by Lee et al. (1997) as well as by Ghodsian (1996). Each of the three to-be-learned sequences had a distinct set of keys to be pressed, was presented in a different color on the computer screen, and had a unique goal MT: 900 ms, 1,200 ms, and 1,500 ms. These MTs were used previously by Lee et al., (1997) as well as by Ghodsian (1996) and were chosen to be well within the response time range of the average person, given some practice.

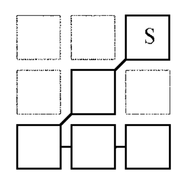
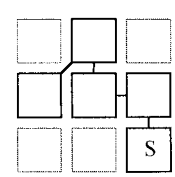
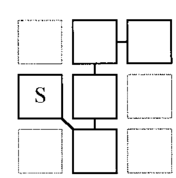
| Keys | Color | Goal MT | Keyboard Pattern |
|---|---|---|---|
| 9-5-1-2-3 | Green | 900 ms |   |
| 3-6-5-8-4 | Red | 1,200 ms |   |
| 4-2-5-8-9 | White | 1,500 ms |   |

*Figure 1.* To-be-learned keystroke sequences. The key labeled "S" was the first in each sequence, and participants had to follow the line to press the subsequent keys in the correct order. MT = movement time; ms = millisecond. Reprinted from "Metacognition in Motor Learning," by D. A. Simon and R. A. Bjork, 2001, *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27,* p. 909. Copyright 2001 by the American Psychological Association.

## Acquisition Phase

*Instructions.* Participants were instructed that the experimental task involved pressing specified sequences of keys on the number pad of the computer keyboard as close as possible to a specified goal MT. They were asked to make all key presses with the index finger of the dominant hand and to make the time between key presses as even as possible, that is, not to start slowly and speed up or vice versa. Keyboard placement was adjusted according to handedness. Participants were also instructed not to press any keys until the pattern to be performed had appeared on the screen and a beep had sounded. The form of feedback was explained, and a printout of the feedback from a sample trial was presented to them, which they then had to interpret verbally to the satisfaction of the experimenter. The nature of the prediction task was also explained, and an example of the prompt that would appear for each prediction was displayed. The instructions stressed the fact that the participants were to estimate how close, in milliseconds, they thought their performance would be to the relevant goal MT if practice were to cease immediately and they were to be given a retention test the next day. Any questions that the participants raised were answered before testing began.

Instructions pertaining to the modeled-timing information were also provided. In the case of the mismatching models, it was important to stress that the models would be different from the trial that would immediately follow. All participants were told that the models would not be present in the retention test the next day. Before the beginning of practice, the participants were encouraged to ask questions about anything that was unclear to them.

*Experimental tasks.* In order to give the participants an opportunity to get used to the task, practice began with five successful trials of a pattern different from the three that were to be learned in the main part of the experiment. Here, and throughout the experiment, for a trial to be deemed successful it was only necessary that the correct keys be pressed in the correct order. Once these five trials of the sample task were completed, the participants were asked if they had any further questions, and then they proceeded with the main acquisition procedures.

The three to-be-learned patterns, as well as their goal MTs and presentation colors, are presented in Figure 1. On each learning trial, the pattern to be performed was presented on the computer screen. The first key of the sequence was marked with an “S” for start, and a line indicated the order in which the remaining keys were to be pressed. As shown in Figure 1, the green pattern required striking the keys 9-5-1-2-3 in a total of 900 ms; the red pattern required striking the keys 3-6-5-8-4 in 1,200 ms; and the white pattern required striking 4-2-5-8-9 in 1,500 ms.

Before each of the first 3 trials of each pattern, the goal MT was presented on the screen for 4 s. Modeled timing information was then presented (see description below). The to-be-performed pattern then appeared on the screen, and the participants were free to respond, as soon as they were ready, by pressing the appropriate keys. The pattern remained visible on the screen until 2 s after the response was completed. After this 2-s delay, feedback was provided about the trial. This feedback, or knowledge of results, consisted of three pieces of information: whether the keys pressed had been correct or wrong; the actual MT for that trial in milliseconds; and the amount, in milliseconds, by which that MT was faster or slower than the goal MT. The feedback remained on the screen for 5 s, and then the screen went blank for 5 s before the next trial began.

The acquisition phase continued, for both blocked- (e.g., 900 ms, 900 ms, 900 ms, . . . , 1,200 ms, 1,200 ms, 1,200 ms . . . , 1,500 ms, 1,500 ms, 1,500 ms . . .) and random-practice conditions (e.g., 900 ms, 1,200 ms, 900 ms, 1,500 ms, 1,200 ms . . . , etc.), until each participant performed 30 successful trials (i.e., made the correct key strokes) for each of the three patterns. If wrong keys were pressed, this fact was indicated in the feedback, and that trial was repeated at the end of the current block. The order in which the patterns were practiced was counterbalanced across participants in both the blocked-practice and random-practice conditions.

*Timing models.* Before each practice trial, all participants saw two computer-generated models. These models consisted of a presentation of the pattern on the computer screen. Beginning with the start key for the pattern, a square corresponding to each key appeared in order on the screen until the whole pattern was visible. A beep occurred coincident with the appearance of each key’s square, and each interkey interval was one quarter of the total goal MT of the pattern. Thus, the whole pattern was “built” both visually and audibly with exactly the timing that would be required for a perfect execution.

In the case of the matching-model conditions, the two models were of the pattern about to be practiced. In the mismatching-model conditions, the two models were the same as each other, but were not models of the pattern about to be executed, rather, they were models of one of the other two experimental patterns. The instructions explicitly stated that the models would not be presented to participants during the test the next day.

*Subjective JOLs.* After every fifth successful trial on each of the three patterns, a question appeared on the computer screen asking the participants to make a judgment about how well they thought they had learned the pattern that they just executed. Specifically, they were asked to predict their own performance on the test scheduled for the next day—assuming that they would get no more practice trials on the pattern in question. The exact wording that appeared on the screen was as follows:

> If practice stopped right now and you were tested tomorrow, . . . . How close do you think you could get to the goal time for the pattern you just performed? Note that on the test you will only see the diagram of the pattern. Type the number of ms and press Enter.

Smaller numbers on this subjective judgment were thus indicative of anticipating more accurate performance on the forthcoming test. As part of the prompt to make a judgment, participants were reminded that on the test to be administered in 24 hrs, they would only see a diagram of the pattern to be performed.

## Retention Phase

On the day after the acquisition phase of the experiment, the participants returned to the experimental room for tests of retention. Before the retention test proper, the participants filled out a brief paper-and-pencil test of their recall of the patterns and associated MTs from the previous day. Three arrays of nine squares (three $\times$ three) corresponding to the nine keys of the number pad were presented on the response sheet. Above each array was the color name corresponding to one of the experimental patterns. The instructions were for each pattern to mark an “s” in the square corresponding to the first key to be pressed for that pattern and then indicate the subsequent keys in their correct order. Beneath each pattern, there was a space in which participants were to write down the recalled MT associated with that particular pattern. Below the space for recalled MT, there was a space for the participant to make a prediction as to their expected performance on each of the patterns. This prediction was in the same form as the predictions they had made throughout the learning phase. That is, they were to write in how far from the target time they thought their retention-test performance would be, in milliseconds. Feedback about their performance on this task was not provided to the participants.

Three trials of each pattern were presented to the participants under a blocked order, and three under a random order, making a total of 18 retention-test trials. The order of blocked and random-test trial presentations was counterbalanced across participants. Both blocked and random orders were used to detect any specificity-of-practice effects (cf. Shea & Morgan, 1979). On these retention-test trials, the diagram of the pattern was presented on the screen— explicit memory of the patterns was therefore not strictly necessary for performance on this test—and the participants responded as soon as they were ready. They were asked to execute a given keystroke pattern in as close as possible to what they remembered the appropriate target time to be. Two seconds after the pattern was executed, a message was displayed on the computer screen informing them whether the keys pressed were correct or wrong. If the wrong keys were pressed, that trial was repeated in a manner concordant with the current test schedule. No information about MTs was provided, either about the goal MTs or as feedback about actual performance. The order of presentation of the patterns within the blocked- and random-test trials for each participant matched the ordering that had occurred in the acquisition phase. Immediately after completion of this performance test, participants filled out brief questionnaires designed to get their impressions of their learning experiences. The questions were (a) "How good was the practice schedule you followed yesterday for learning the movement patterns?" (b) "Overall, how well would you say that you learned the three movement patterns?" (c) "How helpful do you think the computer-generated timing models (i.e., when the computer demonstrated a perfect performance) were for learning the patterns?" Ratings for these questions were made on 6-point Likert-type scales. The remaining two questions were open-ended: (d) "When you made predictions yesterday about your performance today, what did you base those predictions on? Explain briefly"; (e) "If you have any additional comments that you wish to make about the experiment or your practice schedule, please do so here." After completing this questionnaire, the participants were thanked, debriefed, and excused.

## Results

The first section below reports the participants' predictions and actual performance during the acquisition phase. The second section reports their predictions and performance at the time of the final test.

### *Acquisition Phase*

In the analyses of acquisition performance, the trials in the blocked-practice conditions were reordered so as to provide blocks of acquisition trials comparable to those of the random conditions. That is, for both types of schedule, the first 5 correct trials on each pattern were combined to form the first acquisition block; the second 5 correct trials on each pattern were combined to form the second acquisition block; and so on for all the acquisition data.

*Performance measures: %|CE|.* The primary performance measure of interest is absolute constant error (|CE|), a measure reflecting the average deviation, in either direction, of a participant's MTs from the target MTs, that is, response bias with respect to the target.[1] To facilitate comparisons between the different practice patterns, we converted this measure to percent absolute constant error (%|CE|; i.e., by dividing |CE| by the corresponding target MTs and multiplying by 100).

Acquisition %|CE| scores were compared in a 2 × 2 × 6 (Practice Schedule × Model Condition × Trial Block) analysis of variance (ANOVA), where trial block was a within-subjects factor. The overall acquisition data are depicted in the two panels of Figure 2, and the results of the ANOVA are presented in Table 1. Here and for all ANOVA tables, statistically significant effects are accompanied by an effect size measure: Cohen's *f*. For this measure, the commonly used values associated with small, medium, and large effect sizes are 0.1, 0.25, and 0.4, respectively (Cohen, 1988). Main effects emerged for practice schedule and for model condition. Blocked practice yielded lower performance errors than random practice ($M$ = 4.02%, $SD$ = 3.34% vs. $M$ = 6.57%, $SD$ = 5.66%: an effect of medium magnitude), and matching models supported more accurate performance than did mismatching models ($M$ = 3.88%, $SD$ = 2.83% vs. $M$ = 6.71%, $SD$ = 5.87%: an effect of medium magnitude, $f$ = 0.30).

Table 1
*Analysis of Variance for %|CE| in Acquisition*

| Source | *df* | *F* | Cohen's *f* |
|---|---|---|---|
| Between subjects | | | |
| Schedule (Sch.) | 1 | 15.83** | 0.27 |
| Model (M) | 1 | 19.44** | 0.30 |
| Sch. × M | 1 | 1.27 | 0.03 |
| S/Sch. × M (error) | 92 | (59.41) | |
| Within subjects | | | |
| Trial block (T) | | | |
| Linear trends | | | |
| Blocked matching | 1 | 2.60 | 0.04 |
| Blocked mismatching | 1 | 22.60** | 0.15 |
| Random matching | 1 | 13.05** | 0.11 |
| Random mismatching | 1 | 73.50** | 0.28 |
| Contrast error | 92 | (13.33) | |
| Sch. × T | 5 | 3.22** | 0.08 |
| M × T | 5 | 8.89** | 0.15 |
| Sch. × M × T | 5 | 0.46 | — |
| S × T/Sch. × M (error) | 460 | (7.88) | |

*Note.* Values in parentheses represent mean square errors. Dash represents undefined $f$ when the $F$ value is smaller than 1.0 (estimated $\omega^2$ used to calculate $f$). %|CE| = percent absolute constant error; S = subjects.
** $p < .01$.

Linear trend analyses conducted for each condition combination across the six trial blocks of acquisition revealed significant trends for the blocked-mismatching, random-matching, and random-mismatching combinations (see Table 1). However, with the exception of the random-mismatching contrast, the effects were small in magnitude. The linear trend in the blocked-matching combination did not reach conventional levels of significance.

*Predictions: %JOLs.* Predictions about next day performance were made after every five trials of a given pattern during acquisition (see Figure 3). We converted these predictions to percentages of the corresponding target MTs. We took this step to facilitate comparisons to actual performance measures (%|CE|).

We compared predictions made during acquisition in a 2 × 2 × 6 (Practice Schedule × Model Condition × Trial Block) ANOVA, with trial block as a within-subjects factor. The results of this analysis are presented in Table 2. As with actual performance measures, significant main effects emerged for practice schedule and model condition. Blocked practice and matching models supported predictions of more accurate retention performance than did

[1] Absolute constant error (|CE|) is the absolute value of the mean deviation from the target for a set of trials (|CE| = absolute value of: [Σ {$x_i$ − $T$}/$n$], where $x_i$ is the time for Trial $i$, $T$ is the target time, and $n$ is the number of trials in the set; see, e.g., Schmidt & Lee, 1999). Absolute scores are used so that averaging across participants does not lead to an artificially low mean accuracy score. Percent variable errors were also compared, but, other than a main effect in acquisition-model condition such that mismatching models led to greater performance variability than matching models, effects during both acquisition and retention were of limited interest. Hence, presentation of those analyses has been omitted.

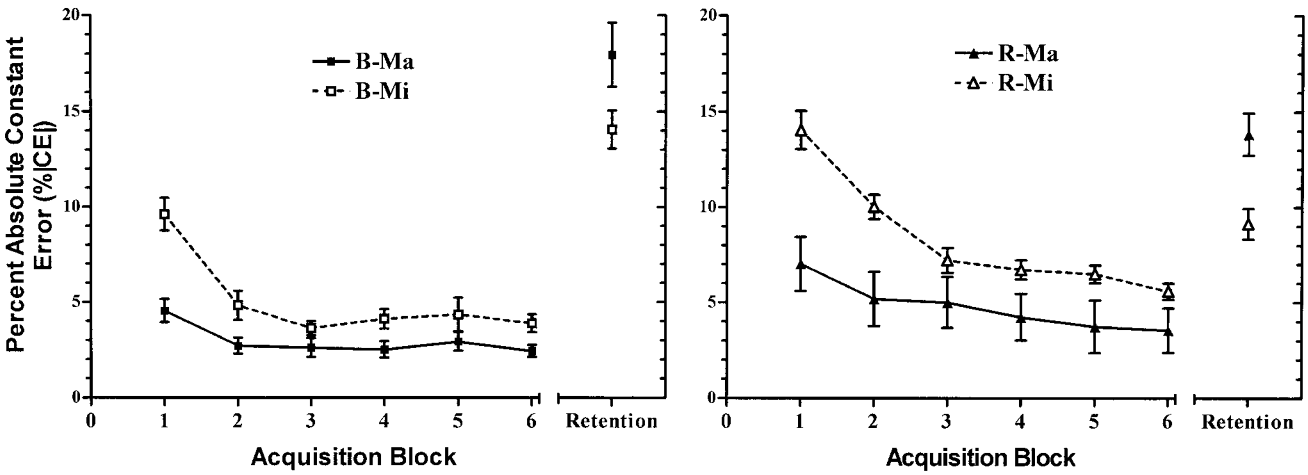



*Figure 2.* Percent absolute constant error (%|CE|) during acquisition and at retention as a function of acquisition condition. Left-hand panel = blocked-matching (B-Ma) and blocked-mismatching (B-Mi) conditions; right-hand panel = random-matching (R-Ma) and random-mismatching (R-Mi) conditions. Error bars represent standard errors of the mean.

random practice and mismatching models, respectively. Both effects were of medium magnitude. The only other significant effect was a main effect of trial block. Participants predicted better future performance after having had more blocks of training on a given pattern. Linear trend analyses conducted across the six blocks of acquisition for the different practice-schedule and model combinations revealed significant results for the blocked-matching, blocked-mismatching, and random-mismatching combinations. However, all of the associated effect sizes were in the range considered small. The linear trend was not statistically significant for the random-matching combination.

### Retention Phase

When participants returned for the retention test, they were first asked to recall both the patterns and target times for the three patterns that they practiced on the previous day. Their responses were prompted by the color name associated with each pattern, and they marked their responses on prepared depictions of the nine keys of the keypad. After attempting to recall the patterns and timing of each pattern, they were asked to predict their performance on each of the patterns in the upcoming retention test. These predictions called for the same kinds of estimates of test accuracy as had been made during practice the previous day.

*Paper-and-pencil tests.* There was a significant main effect of practice schedule on the recall of the keystroke patterns after a 24-hr delay (see ANOVA results in Table 3). The proportion of patterns correctly recalled was lower after a blocked-practice schedule than a random one (*M*s = .326 and .563, *SD*s = .347 and .301, respectively). This effect was medium–large in magni-

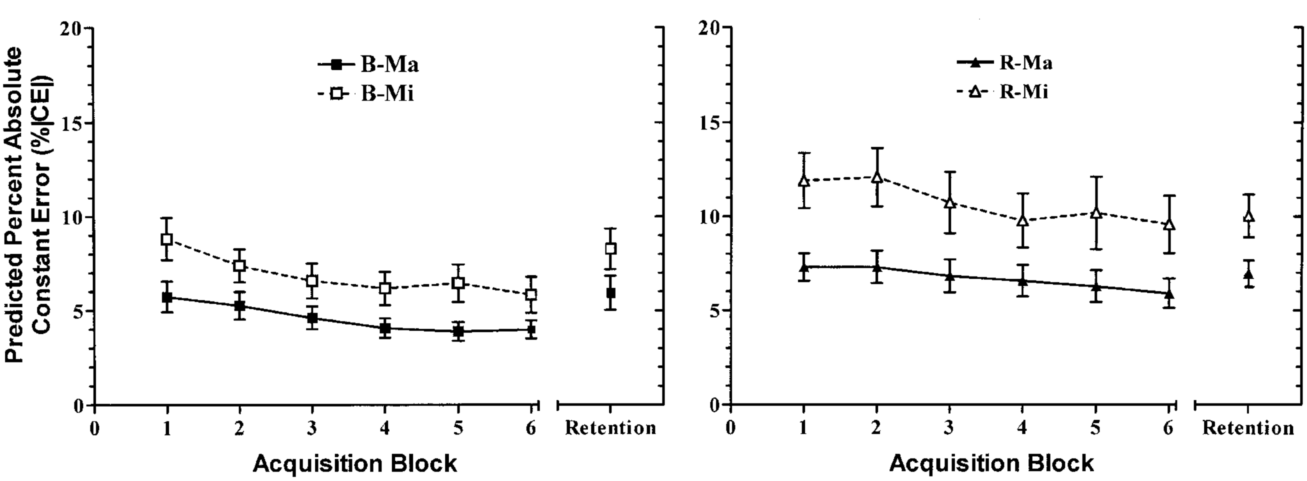



*Figure 3.* Predicted percent absolute constant error (%|CE|) during acquisition and before retention as a function of acquisition condition and trial block. Left-hand panel = blocked-matching (B-Ma) and blocked-mismatching (B-Mi) conditions; right-hand panel = random-matching (R-Ma) and random-mismatching (R-Mi) conditions. Error bars represent standard errors of the mean.

Table 2
*Analysis of Variance for Predicted %|CE| in Acquisition*

| Source | *df* | *F* | Cohen's *f* |
|---|---|---|---|
| Between subjects | | | |
| Schedule (Sch.) | 1 | 9.67** | 0.26 |
| Model (M) | 1 | 10.82** | 0.27 |
| Sch. × M | 1 | 0.82 | — |
| S/Sch. × M (error) | 92 | (131.57) | |
| Within subjects | | | |
| Prediction (P) | | | |
| Linear trends | | | |
| Blocked matching | 1 | 3.97* | 0.05 |
| Blocked mismatching | 1 | 7.24** | 0.07 |
| Random matching | 1 | 2.31 | 0.03 |
| Random mismatching | 1 | 7.43** | 0.07 |
| Contrast error | 92 | (15.56) | |
| Sch. × P | 5 | 0.49 | — |
| M × P | 5 | 0.78 | — |
| Sch. × M × P | 5 | 0.28 | — |
| S × P/Sch. × M (error) | 460 | (6.46) | |

*Note.* Dashes represent undefined *f* when the *F* value is smaller than 1.0 (estimated $\omega^2$ used to calculate *f*). Values in parentheses represent mean square errors. %|CE| = percent absolute constant error; S = subjects.
* $p < .05$. ** $p < .01$.

tude. An interaction between practice schedule and model condition indicated that pattern-recall performance was poorer for those who had practiced under a blocked schedule and with matching models, ($M$ = .19, $SD$ = .29) than for the other three schedule-model combinations, (*M*s = .45, .58, & .54, *SD*s = .35, .33, .27, for blocked-mismatching, random-matching, and random-mismatching conditions, respectively). The effect size was in the lower end of the medium range.

The proportion of correctly recalled target times did not yield a main effect of schedule (see ANOVA results in Table 4), but there was a main effect of model condition such that those who had experienced mismatching models had better recall ($M$ = .694, $SD$ = .356) than those who had matching models ($M$ = .486, $SD$ = .337), and this was a medium-sized effect.

Predictions made about performance immediately before the retention test were analyzed in a 2 × 2 (Practice Schedule × Model Condition) ANOVA (see Figure 3 and Table 5). The main effect for model condition indicated that those who had practiced with matching models were expecting smaller proportional errors ($M$ = 6.45%, $SD$ = 3.98%) than those who had practiced with mismatching models ($M$ = 9.16%, $SD$ = 5.45%). The effect size was of medium magnitude.

*%|CE|.* We analyzed actual performance on the retention trials in a 2 × 2 × 2 (Practice Schedule × Model Condition × Test Schedule) ANOVA, with test schedule as a within-subjects factor (see Figure 2 and Table 6). Main effects of practice schedule and model condition indicated that, contrary both to acquisition performance and predictions made during practice, smaller percent errors were found after random practice ($M$ = 11.5%, $SD$ = 7.1%) than after blocked practice ($M$ = 16.0%, $SD$ = 9.7%), and after mismatching models ($M$ = 11.6%, $SD$ = 6.7%) than after matching models ($M$ = 15.9%, $SD$ = 10.0%). Both of these effects were medium in magnitude.

Given the fact that feedback was not provided to participants concerning their attempts to recall the patterns and their associated target times, it is possible that the observed pattern of %|CE| results is attributable to their misremembering the patterns and then attempting to perform the patterns as close as they could to those incorrect target times. To address this possibility, we performed an analysis in which %|CE| scores for retention performance were determined relative to the remembered MTs instead of the actual target times. In this analysis (see results in Table 7), the only effect to emerge was a main effect of model condition such that having had mismatching models in practice led to smaller retention errors ($M$ = 14.29%, $SD$ = 10.1%) than matching models ($M$ = 21.84%, $SD$ = 17.4%), and this effect was medium in magnitude. However, a comparison of the scores generated in the two different methods indicated that errors around the recalled target times ($M$ = 18.06%, $SD$ = 14.69%) were significantly larger than those around the prescribed MTs ($M$ = 13.75%, $SD$ = 8.75%), $F(1, 92) = 13.56$, $MSE = 131.45$, $p < .05$, Cohen's $f$ = 0.17; an effect in the small–medium range. This latter result

Table 3
*Analysis of Variance for Pattern Recall Scores*

| Source | *df* | *F* | Cohen's *f* |
|---|---|---|---|
| Schedule (Sch.) | 1 | 13.58** | 0.35 |
| Model (M) | 1 | 3.00 | 0.13 |
| Sch. × M | 1 | 5.69* | 0.21 |
| S/Sch. × M (error) | 92 | (0.10) | |

*Note.* Values in parentheses represent mean square errors. S = subjects.
* $p < .05$. ** $p < .01$.

Table 4
*Analysis of Variance for Timing Recall Scores*

| Source | *df* | *F* | Cohen's *f* |
|---|---|---|---|
| Schedule (Sch.) | 1 | 0.62 | — |
| Model (M) | 1 | 8.78** | 0.28 |
| Sch. × M | 1 | 2.50 | 0.12 |
| S/Sch. × M (error) | 92 | (0.12) | |

*Note.* Dash represents undefined *f* when the *F* value is smaller than 1.0 (estimated $\omega^2$ used to calculate *f*). Values in parentheses represent mean square errors. S = subjects.
** $p < .01$.

Table 5
*Analysis of Variance for Preretention Predicted %|CE|*

| Source | *df* | *F* | Cohen's *f* |
|---|---|---|---|
| Schedule (Sch.) | 1 | 2.03 | 0.10 |
| Model (M) | 1 | 7.74** | 0.26 |
| Sch. × M | 1 | 0.14 | — |
| S/Sch. × M (error) | 92 | (22.73) | |

*Note.* Dash represents undefined *f* when the *F* value is smaller than 1.0 (estimated $\omega^2$ used to calculate *f*). Values in parentheses represent mean square errors. %|CE| = percent absolute constant error; S = subjects.
** $p < .01$.

Table 6
*Analysis of Variance for %|CE| in Retention*

| Source | *df* | *F* | Cohen's *f* |
|---|---|---|---|
| Between subjects | | | |
| Schedule (Sch.) | 1 | 8.36** | 0.25 |
| Model (M) | 1 | 7.55** | 0.23 |
| Sch. × M | 1 | 0.06 | — |
| S/Sch. × M (error) | 92 | (117.23) | |
| Within subjects | | | |
| Test (T) | 1 | 0.01 | — |
| Sch. × T | 1 | 0.47 | — |
| M × T | 1 | 0.15 | — |
| Sch. × M × T | 1 | 0.08 | — |
| S × T/Sch. × M (error) | 92 | (21.33) | |

*Note.* Dashes represent undefined *f* when the *F* value is smaller than 1.0 (estimated $\omega^2$ used to calculate *f*). Values in parentheses represent mean square errors. %|CE| = percent absolute constant error; S = subjects.
** $p < .01$.

suggests that although participants may have misremembered the target times, their performance was, nonetheless, more likely centered around the actual target times.

A key question that is not answerable from the present data is how performance in the model conditions compared with the basic blocked- and random-practice conditions. Data from Simon and Bjork (2001) can be brought to bear on this issue. In that study, metacognitive judgments were gathered from participants engaging in blocked or random practice. Participants were drawn from the same population as in the present study. Apart from the absence of timing models, the conditions of practice were identical to those in the present experiment, including common intertrial intervals. Mean retention performance data from the four conditions in the present experiment are presented in Table 8 along with the two conditions from Simon and Bjork (2001). A combined analysis of these data (see results in Table 9) yielded a significant main effect of practice schedule. The effect size was in the medium–large range. There was also a significant main effect of model condition; an effect of small–medium magnitude. Follow-up analyses (Tukey's honestly significant difference test) indicated that there was a difference between the no-model and the mismatching means, and between the matching and mismatching means. There was not a significant interaction between practice schedule and model condition. Thus, overall, the no-model condition was more like the matching than the mismatching-model condition, suggesting that mismatching models support superior retention performance.

*Accuracy of predictions.* To assess the accuracy of the predictions against actual retention performance, we performed a 2 × 2 × 2 (Practice Schedule × Model Condition × Test Schedule) ANOVA on the differences between each participant's retention scores (%|CE| blocked and random test) and their final predictions made during acquisition converted to percentage of target time. Because both the actual (%|CE|) and predicted scores are unsigned estimates of performance, a measure of the difference that is positive indicates that the actual errors were larger than predicted, whereas a negative value indicates that the actual errors were smaller than predicted. Only two significant effects emerged from the analysis of the difference scores: a main effect for practice schedule, and a main effect for model condition (see Table 10). Both of these effects were in the medium–large range. The schedule effect indicated that for the blocked condition, mean differences between predicted and actual errors ($M$ = 11.1%, $SD$ = 10.6%) were bigger than for the random condition ($M$ = 3.7%, $SD$ = 8.4%). The model main effect indicated that difference scores were bigger for those who practiced with matching ($M$ = 10.9%, $SD$ = 10.6%) than for those who practiced with mismatching models ($M$ = 3.9%, $SD$ = 8.6%). Thus, actual performance was closer to predicted and closer to the target values after random than blocked practice, and after mismatching rather than matching models.

*Questionnaire measures.* Interestingly, when asked to rate the "goodness" of their practice schedules (Question 1), there were no significant differences between practice or model conditions (see Table 11 for analyses). However, when asked to indicate how well they felt they had learned the patterns (Question 2), those in the random-practice condition rated themselves significantly higher ($M$ = 4.21, $SD$ = 0.87) than those in the blocked condition ($M$ = 3.63; $SD$ = 1.08); a medium-sized effect. However, there was no significant effect of model condition. When asked about the helpfulness or otherwise of the models during practice for learning the patterns (Question 3), people in the mismatching condition rated the models as less helpful ($M$ = 4.28, $SD$ = 1.32) than did those in the matching condition ($M$ = 5.17, $SD$ = 1.04), a medium- to large-sized effect. However, there was not a significant effect of physical-practice condition. The responses elicited by Questions 4 (bases for making predictions) and 5 (additional comments) yielded responses that were mostly too broad to be informative, for example, "(I based my predictions . . .) on my results yesterday." A few participants provided more useful statements such as, "How close I was on the most recent trials," but there were not enough of either these, or the general comments, to

Table 7
*Analysis of Variance for Retention %|CE| Based on Recalled MTs*

| Source | *df* | *F* | Cohen's *f* |
|---|---|---|---|
| Between subjects | | | |
| Schedule (Sch.) | 1 | 0.01 | — |
| Model (M) | 1 | 6.99** | 0.24 |
| Sch. × M | 1 | 0.61 | — |
| S/Sch. × M (error) | 92 | (391.26) | |
| Within subjects | | | |
| Test (T) | 1 | 0.08 | — |
| Sch. × T | 1 | 0.03 | — |
| M × T | 1 | 1.21 | 0.01 |
| Sch. × M × T | 1 | 2.02 | 0.02 |
| S × T/Sch. × M (error) | 92 | (23.64) | |

*Note.* Dashes represent undefined *f* when the *F* value is smaller than 1.0 (estimated $\omega^2$ used to calculate *f*). Values in parentheses represent mean square errors. %|CE| = percent absolute constant error; MTs = movement times; S = subjects.
** $p < .01$.

Table 8
*Mean Retention Performance (%|CE|) as a Function of Acquisition Practice Schedule and Modeled-Timing Condition*

| Practice schedule | Model condition: No model[a] | Matching | Mismatching | Average |
|---|---|---|---|---|
| Blocked | 19.68 (8.39) | 17.96 (11.53) | 14.06 (6.93) | 17.23 (9.39) |
| Random | 10.80 (5.97) | 13.84 (7.72) | 9.15 (5.57) | 11.26 (6.73) |
| Average | 15.24 (8.51) | 15.90 (9.97) | 11.61 (6.72) | |

*Note.* Values in parentheses are standard deviations. %|CE| = percent absolute constant error.
[a] Data from Simon and Bjork (2001).

perform any serious analysis of responses to these questionnaire items.

## Discussion

The key performance results of the present study are that during acquisition, blocked practice supported lower errors in MT than random practice, and matching models also supported lower errors than mismatching models—both medium-sized effects, so the impact of the schedule and modeling conditions on the practiced tasks was meaningful—but there was no interaction between physical practice and model conditions. In retention however, these relationships were reversed; the random-practice condition led to lower errors than blocked practice, and mismatching models during practice supported more accurate performance than matching models; once more, these were both medium-sized effects. Again, there was not a significant interaction between practice schedules and model conditions. The contextual interference effect (e.g., Shea & Morgan, 1979) was thus replicated, and the novel-practice intervention of providing mismatching models was shown to improve performance relative to the matching-model condition.

Metacognitive judgments of learning made during practice and immediately before retention trials were at odds with actual performance on those trials. It appears that JOLs were made on the basis of immediate performance during practice, rather than being based on any accurate sense of what learning was actually occurring during practice. This effect is particularly noteworthy in the preretention judgments: It is perhaps not surprising that immediate performance considerations should dominate metacognitions during practice, but it is more surprising that at a delay, an entire day removed from the practice phase, learners did not have stronger insight into their immediate ability to perform the experimental tasks. This fact is even more surprising given that immediately before making these preretention JOLs, incidental recall of the patterns was poorer for those in the blocked condition, and incidental recall of the goal MTs was poorer in the matching-model conditions. Rather counterintuitively, the relatively poor recollection of the pattern information apparently did not undermine these participants' performance expectations. The medium effect size measures for these main effects suggest that they were meaningful with respect to the influence of the manipulated variables on predictions made by the participants.

Predictions about delayed performance are important, particularly where learning is a self-paced process, because if the practice situation conveys the impression that learning is more developed than it actually is, learners would seem more likely to terminate their training earlier rather than later. If the practice situation provides a false sense of mastery, learners would presumably spend less time on the task rather than more. It should be noted that in the standard contextual interference effect, superior learning emerges from random rather than from blocked practice when the number of training trials is held constant across the two schedules. It seems likely then that the learning disadvantage for blocked practice (or matching models) would only be amplified if learners were to use immediate performance as their guide to progress in learning and, thus, in assessing how much more they need to practice.

In terms of retention performance, blocked-mismatching and the random-matching conditions performed very similarly (%|CE| *M*s = 14.1% and 13.8%, respectively; see Figure 1). One may argue that each of these groups was subject to one form of beneficial interference and that they were somewhat equivalent in their impact on learning. Having said this, however, the presence of main effects favoring blocked practice and matching models in acquisition, and random practice and mismatching models in retention, coupled with a lack of significant interaction between practice schedule and matching condition at both stages, suggests

Table 9
*Analysis of Variance of Retention %|CE|, Including Data From Simon and Bjork (2001)*

| Source | *df* | *F* | Cohen's *f* |
|---|---|---|---|
| Between subjects | | | |
| Schedule (Sch.) | 1 | 24.12** | 0.36 |
| Model (M) | 2 | 4.82** | 0.20 |
| Sch. × M | 2 | 1.47 | 0.06 |
| S/Sch. × M (error) | 138 | (106.45) | |
| Within subjects | | | |
| Test (T) | 1 | 0.04 | — |
| Sch. × T | 1 | 0.35 | — |
| M × T | 1 | 0.16 | — |
| Sch. × M × T | 1 | 0.10 | — |
| S × T/Sch. × M (error) | 138 | (22.00) | |

*Note.* Dashes represent undefined *f* when the *F* value is smaller than 1.0 (est. $\omega^2$ used to calculate *f*). Values in parentheses represent mean square errors. %|CE| = percent absolute constant error; S = subjects.
** $p < .01$.

Table 10
*Analysis of Variance for Accuracy of Predictions*

| Source | *df* | *F* | Cohen's *f* |
|---|---|---|---|
| Between subjects | | | |
| Schedule (Sch.) | 1 | 18.36** | 0.37 |
| Model (M) | 1 | 16.91** | 0.35 |
| Sch. × M | 1 | 0.58 | — |
| S/Sch. × M (error) | 92 | (141.05) | |
| Within subjects | | | |
| Test (T) | 1 | 0.01 | — |
| Sch. × T | 1 | 0.47 | — |
| M × T | 1 | 0.15 | — |
| Sch. × M × T | 1 | 0.08 | — |
| S × T/Sch. × M (error) | 92 | (21.33) | |

*Note.* Dashes represent undefined *f* when the *F* value is smaller than 1.0 (estimated $\omega^2$ used to calculated *f*). Values in parentheses represent mean square errors. S = subjects.
** $p < .01$.

an additive effect for the two manipulations of the practice experience. This apparent additivity in turn suggests that the effects may be independent of one another, and may tap into separate processes to affect performance and learning. Although the pattern of results per se is not proof of such independence, it does at least bring into question the initial interpretation of Lee et al.'s (1997) findings, that is, that their models were disrupting the very same processing that would normally occur in random practice. As such, a reconsideration of their results as evidence for the reconstruction hypothesis may be warranted. Clearly, this issue merits further exploration.

From an applied perspective, the utility of models such as the ones used here is, we think, quite straightforward. Often training situations occur where a delay is necessitated between practice trials, such as when there is an equipment-to-trainee ratio less than 1:1, or when the physiological demands of the practiced skill require a brief recovery period—both would be common in activities such as gymnastics, trampolining, and diving, among others. The model data lead us to suggest that there would be a greater learning benefit if during these enforced delays, trainees watched another person execute a skill different from the one they were about to perform themselves rather than that same skill, and especially so when they were engaging in blocked practice.

Findings from a study by Gabriele, Hall, and Lee (1989) are particularly interesting in light of the current findings. In their experiment, participants were selected on the basis of their mental imagery abilities, and then they physically practiced four different barrier-knockdown patterns to a criterion of performance. This practice was conducted under either blocked or random conditions. Between trials of physical practice, the learners were asked to mentally image practice of the patterns in either a blocked or a random fashion. For the blocked-imagery trials, the pattern being imaged was the same as the one just executed physically and was imaged three times. For the random intervening imagery, the three patterns other than the one just physically practiced were imaged after each physical-practice trial. Notably, random physical practice led to significantly better retention performance than blocked physical practice, and random imagery led to significantly better retention performance than blocked imagery. In addition, there was no significant interaction between the physical- and imagery-schedule manipulations. The similarity of outcomes between these data and those of the present study are quite striking, and they both serve to underline the beneficial role of interleaving both physical and cognitive processing as an aid to learning.

The similarity between the outcomes of the Gabriele et al. (1989) study and the present one highlights an important issue that has yet to be clearly understood: the nature of the interleaved activity between learning trials that is necessary to benefit learning. Clearly, as shown here and by Lee et al. (1997), matching models do not benefit learning, but what activity other than mismatching models, or mental imagery of the other to-be-learned tasks, would act to provide enough beneficial interference to help learning? This question is not new (see, e.g., Lee, 1988), and a clear answer to it would likely go a long way toward settling the question of why contextual interference works, and which of the theoretical accounts proposed to date are most accurate. Further studies should address this issue.

## Conclusion

The present findings, along with other similar findings in the literature, have clear implications for how individuals responsible for training should structure the conditions of practice. The metacognitive data from the present study, however, along with those we reported earlier (Simon & Bjork, 2001), suggest that students and trainees may be less accurate in appraising their own level of skill learning if their teacher or trainer structures the conditions of teaching and training in ways that optimize long-term retention

Table 11
*Analyses of Variance for Questionnaire Items*

| | Question 1 | | | Question 2 | | | Question 3 | | |
|---|---|---|---|---|---|---|---|---|---|
| Source | *df* | *F* | Cohen's *f* | *df* | *F* | Cohen's *f* | *df* | *F* | Cohen's *f* |
| Schedule | 1 | 1.82 | 0.09 | 1 | 8.53** | 0.28 | 1 | 2.53 | 0.12 |
| Model | 1 | 0.38 | — | 1 | 0.38 | — | 1 | 13.37** | 0.35 |
| Schedule × Model | 1 | 1.22 | 0.11 | 1 | 1.22 | 0.05 | 1 | 0.05 | — |
| Error | 92 | (0.69) | | 92 | (0.96) | | 92 | (1.41) | |

*Note.* Values in parentheses represent mean square errors. Dashes represent undefined *f* when the *F* value is smaller than 1.0 (estimated $\omega^2$ used to calculate *f*).
** $p < .01$.

and transfer. Our findings suggest that learners tend to base their assessments of how much they have learned on their immediate performance, and that their judgments about future performance are quite inaccurate. That is, they are apparently unable to dissociate immediate from longer-term changes in their ability to perform the to-be-learned tasks. Toward the goal of optimizing instruction and training, instructors and learners both need to become aware that immediate performance is an unreliable guide to the long-term changes that are the goal of instruction and training.

## References


Battig, W. F. (1972). Intratask interference as a source of facilitation in transfer and retention. In R. F. Thompson & J. F. Voss (Eds.), *Topics in learning and performance* (pp. 131–159). New York: Academic Press.

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.

Benjamin, A. S., Bjork, R. A., & Schwartz, B. L. (1998). The mismeasure of memory: When retrieval fluency is misleading as a metamnemonic index. *Journal of Experimental Psychology: General, 127,* 55–68.

Bjork, R. A. (1999). Assessing our own competence: Heuristics and illusions. In D. Gopher & A. Koriat (Eds.), *Attention and performance XVII. Cognitive regulation of performance: Interaction of theory and application* (pp. 435–459). Cambridge, MA: MIT Press.

Boyce, B. A., & Del Rey, P. (1990). Designing applied research in a naturalistic setting using a contextual interference paradigm. *Journal of Human Movement Studies, 18,* 189–200.

Carlson, R. A., Sullivan, M. A., & Schneider, W. (1989). Practice and working memory effects in building procedural skill. *Journal of Experimental Psychology: Leaning, Memory, and Cognition, 15,* 517–526.

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

Chamberlin, C., & Lee, T. D. (1993). Arranging practice conditions and designing instruction. In R. N. Singer, M. Murphey, & L. K. Tennant (Eds.), *Handbook of research on sport psychology* (pp. 213–241). New York: Macmillan.

Christina, R. W., & Bjork, R. A. (1991). Optimizing long-term retention transfer. In D. Druckman & R. A. Bjork (Eds.), *In the mind's eye: Enhancing human performance* (pp. 23–56). Washington, DC: National Academy Press.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Gabriele, T. E., Hall, C. R., & Lee, T. D. (1989). Cognition in motor learning: Imagery effects on contextual interference. *Human Movement Science, 8,* 227–245.

Ghodsian, D. (1996). *Training-based predictors of long-term retention and transfer* (Doctoral dissertation, University of California, Los Angeles, 1996). *Dissertation Abstracts International 57(6-B),* 4049.

Ghodsian, D., Bjork, R. A., & Benjamin, A. S. (1997). Evaluating training during training: Obstacles and opportunities. In M. A. Quiñones & A. Ehrenstein (Eds.), *Training for a rapidly changing workplace: Applications of psychological research* (pp. 63–98). Washington, DC: American Psychological Association.

Goode, S., & Magill, R. A. (1986). Contextual interference effects in learning three badminton serves. *Research Quarterly for Exercise and Sport, 57,* 308–314.

Hall, K. G., Domingues, D. A., & Cavazos, R. (1994). Contextual interference effects with skilled baseball players. *Perceptual and Motor Skills, 78,* 835–841.

Lee, T. D. (1988). Transfer-appropriate processing: A framework for conceptualizing practice effects in motor learning. In O. G. Meijer & K. Roth (Eds.), *Complex motor behaviour: The motor-action controversy* (pp. 201–215). Amsterdam: Elsevier.

Lee, T. D., & Magill, R. A. (1983). The locus of contextual interference in motor-skill acquisition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 9,* 730–746.

Lee, T. D., & Magill, R. A. (1985). Can forgetting facilitate skill acquisition? In D. Goodman, R. B. Wilberg, & I. M. Franks (Eds.), *Differing perspectives in motor learning, memory, and control* (pp. 3–22). Amsterdam: North-Holland.

Lee, T. D., Wishart, L., Cunningham, S., & Carnahan, H. (1997). Modeled timing information during random practice eliminates the contextual interference effect. *Research Quarterly for Exercise and Sport, 68,* 100–105.

Magill, R. A., & Hall, K. G. (1990). A review of the contextual interference effect in motor skill acquisition. *Human Movement Science, 9,* 241–289.

McCullagh, P. (1993). Modeling: Learning, developmental and social psychological considerations. In R. N. Singer, M. Murphey, & L. K. Tennant (Eds.), *Handbook of research on sport psychology* (pp. 106–126). New York: Macmillan.

Metcalfe, J., & Shimamura, A. P. (Eds.). (1994). *Metacognition: Knowing about knowing.* Cambridge, MA: MIT Press.

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.

Schmidt, R. A., & Lee, T. D. (1999). *Motor control and learning: A behavioral emphasis* (3rd ed.). Champaign, IL: Human Kinetics.

Schmidt, R. A., Young, D. E., Swinnen, S., & Shapiro, D. C. (1989). Summary knowledge of results for skill acquisition: Support for the guidance hypothesis. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15,* 352–359.

Shea, J. B., & Morgan, R. L. (1979). Contextual interference effects on the acquisition, retention, and transfer of a motor skill. *Journal of Experimental Psychology: Human Learning and Memory, 5,* 179–187.

Shea, J. B., & Zimny, S. T. (1983). Context effects in memory and learning movement information. In R. A. Magill (Ed.), *Memory and control of action* (pp. 345–366). Amsterdam: Elsevier.

Shea, J. B., & Zimny, S. T. (1988). Knowledge incorporation in motor representation. In O. G. Meijer & K. Roth (Eds.), *Complex movement behaviour: The motor-action controversy* (pp. 289–314). Amsterdam: Elsevier.

Simon, D. A., & Bjork, R. A. (2001). Metacognition in motor learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27,* 907–912.



Received May 31, 2001
Revision received June 11, 2002
Accepted June 12, 2002 ■