Journal of Experimental Psychology:
Learning, Memory, and Cognition
2010, Vol. 36, No. 5, 1134–1149

0278-7393/10/$12.00 DOI: 10.1037/a0019998

Loss of Cognitive Skill Across Delays: Constraints for Theories of Cognitive Skill Acquisition

Nicolas J. Wilkins and Katherine A. Rawson
Kent State University

Mastering a cognitive skill requires many practice sessions, occurring over a period of days, weeks, months, or even years. Although a large body of research describes and explains gains made within a given practice session, few studies have investigated what happens to these gains across a delay, and none have examined effects of delays on item-general gains. Across 3 experiments, participants performed alphabet arithmetic verification in an initial practice session followed by a test session after a delay (from 0 to 30 days). All experiments included conditions yielding item-general practice gains; Experiments 2–3 also included an item-specific practice condition. Surprisingly, item-general gains were relatively well preserved across a delay (e.g., only 6.7% decrease in practice effects after 2 days), whereas item-specific gains showed sizeable losses across a delay (e.g., 25.9% loss after 2 days). Results provide important empirical constraints to theories of cognitive skill acquisition.

Keywords: skill acquisition, automaticity, forgetting, potentiation, practice effects

Practice effects are among the most robust findings in cognitive psychology, with a wealth of research showing speed and/or accuracy gains during practice in various cognitive tasks with various kinds of learners (e.g., Logan, 1988; McAndrews & Moscovitch, 1990; Rawson, 2004; Rawson & Middleton, 2009; Rawson & Touron, 2009; Schneider & Shiffrin, 1977; Touron & Hertzog, 2004; Touron, Swaim, & Hertzog, 2007). Most research on cognitive skill acquisition has focused on what types of gains are made during practice and the mechanisms that underlie gains made during practice. However, acquiring a cognitive skill requires many practice sessions, which may be separated by delays of days, weeks, months, or even years. Surprisingly, the effects of a delay on practice gains has largely been ignored in this literature (in stark contrast to the extensive emphasis on quantifying and explaining forgetting in the literature on explicit/episodic memory; see Rubin, Hinton, & Wenzel, 1999; Rubin & Wenzel, 1996; Wixted & Carpenter, 2007).

Given that the literature on cognitive skill acquisition has largely ignored the effects of delays on practice gains, systematic empirical research is clearly needed. If significant changes in practice gains occur between sessions, "a complete theory of human skill acquisition must therefore account for the effects of the delays between sessions on learning and performance" (Rickard, 2007, p. 297). An important question thus arises: What are the effects of a delay on practice gains?

Accordingly, the goal of the current research was to explore the effects of a delay on gains made during a practice session. Such an investigation must take into account the different types of practice gains, which may be affected differently by a delay between practice sessions. Consequently, we first briefly discuss the different types of gains that can occur during practice. Second, we provide a brief overview of the few studies that have investigated the effects of a delay on practice gains and what these studies suggest about the effects of a delay on practice gains. Finally, we present three new experiments extending beyond previous research.

Types of Gains During Practice

Although practice can improve both accuracy and speed of task performance, most research has focused on improvements in speed (e.g., Bajic & Rickard, 2009; Gaschler & Frensch, 2007; Logan, 1992; Logan & Etherton, 1994; Rawson & Middleton, 2009). Indeed, a negatively accelerating decrease in speed of responding with practice is generally considered to be the signature pattern of automatization, and theories of human skill acquisition must account for this speedup to be viable. Complete theories also must account for the different types of practice gains that occur. Practice gains come in two general forms—item-general gains and item-specific gains (defined next)—and different underlying mechanisms have been proposed to account for these two types of gains.

Item-general gains are defined as speedups with practice that accrue to all stimulus tokens of a given type, including both practiced and novel tokens of that type. To the extent that practice gains transfer to novel tokens, practice gains are item-general. Item-general gains have been accounted for by several different mechanisms, most of which involve improvements in the efficiency of algorithmic processes that interpret tokens of a given

Nicolas J. Wilkins and Katherine A. Rawson, Department of Psychology, Kent State University.

This research was partially supported by a Collaborative Award from the James S. McDonnell Foundation 21st Century Science Initiative in Bridging Brain, Mind and Behavior. Thanks to Michael Appleman, Nicole Gonzalez, Phil Grimaldi, Caitlin Metelko, Dan Molnar, Rochelle O'Neil, Sara Smith, and Kathryn Wissman for assistance with data collection.

Correspondence concerning this article should be addressed to Nicolas J. Wilkins, Department of Psychology, Kent State University, P.O. Box 5190, Kent, OH 44242-0001. E-mail: nwilkins@kent.edu

type. For example, according to adaptive control of thought–rational (ACT-R; Anderson & Lebiere, 1998), an algorithm consists of a sequence of productions that fire in serial fashion to compute a response for a stimulus. At each step in the sequence, a production must be selected from among competitors (all productions that satisfy the current goal). With practice, productions that lead to a correct response accrue strength, and strength influences both the likelihood of selection and the speed of the production. Thus, the efficiency of the algorithm improves, which benefits any token of the type processed by that algorithm.

Item-specific gains are defined as speedups with practice that accrue only to the particular tokens that have been practiced and not to novel tokens of the same type. To the extent that practice gains are greater for practiced versus novel tokens, practice gains are item-specific. Item-specific practice gains have been explained by memory-based theories of automaticity (Logan, 1988; Palmeri, 1997; Rickard, 1997). Although these theories differ on some specific assumptions, they share core assumptions relevant to explaining item-specific practice gains. All memory-based theories assume that there are two routes for responding to stimuli: an algorithmic route and a memory-retrieval route. On the first encounter with a specific stimulus, the item-general algorithmic route generates an appropriate response for the stimulus, and a memory trace for the event is encoded into memory. Upon subsequent encounters with the stimulus, processing may again be based on the item-general algorithmic route or it may be based on retrieving the previously stored trace from memory. As the number or strength (depending on the particular memory-based account) of memory traces increases with practice, retrieval speed increases. As a result, the retrieval route can output an interpretation more quickly than the item-general algorithm and thus becomes the primary route used to respond to a repeated stimulus. The memory-retrieval route is necessarily item-specific because in order to retrieve a memory trace the stimulus must have been previously processed to have been encoded.

Prior Research on the Effect of Delays on Practice Gains

Relative to the amount of research on the types of gains within a practice session and their underlying mechanisms, minimal research has directly investigated what happens to practice gains between practice sessions. In one recent study, Rickard (2007) presented five data sets from four separate studies involving various tasks (multiplication and division, digit entry, pound arithmetic, and letter–digit paired-associate learning). Each task shared the common method of repeatedly presenting a set of stimuli to participants during multiple practice sessions, each session separated by 2 days. For the present purposes, we focus here on data from the pound arithmetic task and the paired-associate task and on the delay between Session 1 and Session 2.

Of primary interest, Rickard (2007) found that participants' response times were slower at the beginning of Session 2 than at the end of Session 1, indicating some loss of practice gains across a delay. To quantify the relative amount of loss across the delay, we calculated the percent of Session 1 gains lost at the beginning of Session 2 (hereafter referred to as *percentage of gain lost*, or PGL).[1] To compute the PGL, we first calculated the gain made during Session 1 by subtracting the mean response time for the last block of practice from the mean response time for the first block of practice. We then calculated the loss across the delay between practice sessions by subtracting the mean response time for the last block of Session 1 from the mean response time for the first block of Session 2. Finally, we computed the PGL as (loss/gain) $\times$ 100. As reported in Table 1, the PGL was 19% for the pound arithmetic task and 41% for the paired-associate task.

Results of the three other studies summarized in Table 1 also reveal loss of practice gains across a delay, suggesting that complete theories of cognitive skill acquisition will need to account for what happens to practice gains across a delay. Importantly, the degree of loss of practice gains across a delay varied widely across studies, with a minimum PGL of 13% after a delay of over a year (Kolers, 1976) to a maximum PGL of 83% after a delay of a week (Anderson, Fincham, & Douglass, 1999, Experiment 2).

What explains the large variability of PGL values across the different tasks? One possibility is that the practice gains accrued in the various tasks differentially involved item-general and item-specific practice gains. Although each study involved repetition of specific items during practice, participants' gains for some of these tasks may nonetheless have involved a mix of item-general and item-specific practice gains. For example, pound arithmetic (Rickard, 1997, 2007; Rickard & Bourne, 1995) involves solving condensed algorithmic equations (e.g., 4 # 17 = ?). To solve the problems, participants first subtract the left number from the right number (17 – 4 = 13), add one to the sum (13 + 1 = 14), and then add this sum to the right-most number presented in the problem (17 + 14 = 31). In terms of ACT-R (Anderson & Lebiere, 1998), practice effects (particularly early in practice) may have been due in part to improvements in the likelihood and speed of selecting appropriate productions (item-general algorithm efficiency gains), whereas practice effects later in practice may have been due in part to directly retrieving intermediate or final solutions from memory (item-specific practice gains). Indeed, participants' strategy reports suggest that response time gains during Session 1 were mostly due to item-general gains (Rickard, 1997). At the end of Session 1 (the ninth block of practice), an estimated 85% of the trials were solved by use of an algorithm (estimate derived from Rickard, 1997, Figure 7).

To what extent does the type of practice gain account for the range of PGL values estimated in other studies? Rickard, Lau, and Pashler (2008) reported two experiments involving 15 repetitions of multiplication problems during Session 1. Problems were repeated with a lag of either 2 or 11 other items between trials. Problems were presented again in a second session either 7 days later (Experiment 1) or 2 days later (Experiment 2). On the basis of estimates from data reported in the figures of Rickard et al.'s article, PGL values were larger for the lag-2 condition than for the lag-11 condition. Rickard et al. had participants report the strategy used to solve problems (computation or retrieval) for the last five blocks of practice during Session 2 of Experiment 2. Participants reported greater use of retrieval in the lag-2 condition (>80%) than in the lag-11 (50%–60%) condition. To the extent that retrieval-based processing yields item-specific practice gains, these results suggest that practice gains were largely item-specific

[1] Thanks to Timothy C. Rickard for providing us with the data from these studies.

Table 1
Summary of Data From Previous Research

Study	Test delay	S1 start RT	S1 end RT	S2 start RT	Gain	Loss	PGL	Type of practice gain
Rickard (2007)[a]								
Pound arithmetic	2 days	12,386	5,424	6,720	6,962	1,296	19.0	IS/IG
Letter–digit pair learning	2 days	1,475	701	1,022	774	321	41.0	IS
Rickard, Lau, & Pashler (2008)[a]								
Experiment 1: multiplication, lag 2	7 days	5,300	1,100	4,150	3,900	3,050	78.2	IS
Experiment 1: multiplication, lag 11	7 days	4,900	2,650	3,350	2,250	700	31.1	IS/IG
Experiment 2: multiplication, lag 2	2 days	4,400	1,400	3,700	3,000	2,300	76.7	IS
Experiment 2: multiplication, lag 11	2 days	4,700	2,700	3,000	2,000	300	15.0	IS/IG
Anderson, Fincham, & Douglass (1999)[b]								
Experiment 1: rule application	1 day	37.3	10.5	14.1	26.8	3.6	13.4	IS/IG
	7 days	38.7	10.3	17.5	28.4	7.2	25.4	IS/IG
Experiment 2: fact retrieval	1 day	13.6	7.1	9.2	6.5	2.1	32.3	IS
	7 days	14.1	6.9	12.9	7.2	6.0	83.3	IS
Kolers (1976)[c]								
Inverted text	13–15 months	15.8	2.0	3.8	13.8	1.8	13.0	IS/IG

Note. S1 = Session 1; RT = response time; S2 = Session 2; PGL = percentage of gain lost ([loss/gain] × 100); IS = item-specific; IG = item-general.
[a] Mean RTs are in milliseconds for S1 start RT, S1 end RT, S2 start RT, Gain, and Loss columns. [b] Mean RTs are in seconds for S1 start RT, S1 end RT, S2 start RT, Gain, and Loss columns. [c] Mean RTs are in minutes for S1 start RT, S1 end RT, S2 start RT, Gain, and Loss columns.

in the lag-2 condition and a mix of item-general and item-specific gains in the lag-11 condition.

In Experiment 1 of Anderson et al., (1999), participants first learned eight associative facts for when a particular event occurred on two separate days (e.g., Skydiving: Wednesday at 2:00 and Thursday at 12:00). Learning of the eight associative facts consisted of an initial presentation phase followed by three dropout blocks of practice in which items were learned to a criterion of one correct recall. After learning the eight facts, participants then performed a rule application task for four sessions. The rule application task involved presenting participants with novel stimuli—each consisting of an event plus one day and one time that the event occurred—and blanks representing the next day and time the event occurred (e.g., Skydiving: Tuesday at 3:00 and ______ at _____). To respond to the novel stimuli, participants retrieved the previously learned associative facts in order to use the rules of association to calculate the missing day and time (the second skydiving was 1 day later and 2 hr earlier than the first). Of interest here, the PGL across the delay between Sessions 1 and 2 were calculated on the basis of data taken from the ACT-R web page (http://act-r.psy.cmu.edu/files/oldmodels/practice-and-retention), and moderate loss of practice gains were found (see Table 1). Unlike the other tasks discussed earlier, the rule application task did not repeat specific items during practice, which might suggest that gains made during practice were item-general. However, the algorithm used by participants to solve stimuli in the rule application task required retrieval of previously learned associative facts. Given that specific associative facts were repeatedly retrieved during the task, part of the gain made during practice was likely item-specific.

In Anderson et al.'s (1999) Experiment 2, participants learned the same eight associative facts as in Experiment 1 but then practiced a fact retrieval task. Each trial consisted of repeatedly presenting one half of a previously learned associative fact, and participants had to retrieve the other half of the associated day and time. Calculation of the PGL from the data indicated substantial loss of practice gains across the delays (see Table 1).

Finally, Kolers (1976) had participants read 160 unique pages of typographically inverted text during practice. After completion of the practice session, participants then returned 13–15 months later for a second session in which they read both previously practiced pages of inverted text and novel pages of inverted text. On the basis of estimates from data reported in figures of Kolers' article, our PGL calculation indicated a modest loss of practice gains across the delay (see Table 1). Kolers reported two empirical findings suggesting that gains made during practice were a mix of item-general and item-specific practice. Concerning item-specific gains, during Session 2, mean reading times were 5% slower for novel pages than for practiced pages. Concerning item-general gains, reading time per page was faster during Session 2 than at the start of Session 1 for both repeated and novel pages of text. However, what appear to be item-general gains at the page level may have been partly due to item-specific gains at the word level. Although unique pages were used as either repeated materials or novel materials, at least some words likely repeated across pages. At a minimum, an examination of reading times at the level of the word would be needed to confidently diagnose item-general gains versus item-specific gains for repeated token words.

If our task analysis is correct, then a potential explanation emerges for the large range of PGL values across studies: The PGL was consistently lower for tasks with a mix of item-general and item-specific practice gains than for tasks with primarily item-specific practice gains (range of 13%–31% vs. 32%–83%, respectively). Lower PGL for tasks with a mix of item-specific and item-general practice gains suggests that item-general gains may be more resistant to loss across a delay than are item-specific practice gains.

However, in none of the studies discussed previously were Session 1 practice gains clearly item-general effects. Thus, a major goal of the current research was to examine the effects of a delay on item-general practice gains. Additionally, given that no prior research systematically examined the effects of a delay as a function of type of gain, the current research also compared the effects of a delay on item-general practice gains with those on item-specific gains.

To revisit our purpose, our highest level question concerns the effect of a delay on practice gains. Thus far, we have focused our review of previous literature on the loss of practice gains across a delay. However, Rickard (2007) reported one other observed effect of a delay on practice gains, which he referred to as *potentiation*. Potentiation is the empirical finding that response times during a second session of practice are faster than what would be predicted by extrapolating from response times in Session 1 through the trials of Session 2. Potentiation suggests that a delay may somehow facilitate future learning. However, unlike in Rickard's studies, the data reported by Anderson et al. (1999) and Kolers (1976) do not show clear signs of potentiation. Although potentiation is of secondary interest here, we test for this effect in the current research because of the mixed findings from previous research.

Current Research

The primary goals of the current research were to examine the effects of various delays on item-general practice gains (Experiment 1) and to directly compare the loss of item-general practice gains with the loss of item-specific practice gains (Experiments 2 and 3). To directly examine these effects, in all experiments we used an alphabet arithmetic (AA) verification task (Compton & Logan, 1991; Logan, 1988; Logan & Klapp, 1991). In the AA verification task, participants verify whether AA problems (e.g., A + 2 = C) are true or false as quickly as possible. Participants in both experiments performed a practice session and a test session separated by a delay. To diagnose type of practice gain, we also included novel AA items in the test session.

As an additional means to diagnose type of gain, the addend size of the AA problems was varied. Increasing the addend size of AA problems presumably increases the number of algorithmic steps of counting up the alphabet. Thus, if participants' responses are computed using an algorithm, response times will increase monotonically with addend size. In contrast, memory-based theories assume that item-specific gains made during practice reflect directly retrieving a response from memory. Prior research has shown that when participants shift to retrieving answers from memory, response times are similar for items of all addend sizes because retrieval is a one-step process (e.g., Logan, 1988). Thus, significant differences in response times as a function of addend size at the end of practice would suggest that items are still being processed via algorithm, which produces item-general gains. No differences in response times as a function of addend size at the end of practice suggests that items are being processed via retrieval, which produces item-specific gains.

Experiment 1 was designed to study the effects of a delay on item-general practice gains. The primary manipulation was delay (0, 1, 7, or 30 days). To foreshadow, preservation of item-general practice gains was surprisingly good relative to the loss of item-specific practice gains in previous research discussed earlier. Experiments 2 and 3 were designed to more directly compare loss of item-general practice gains with loss of item-specific practice gains across a delay.

Experiment 1

Method

Participants and design. Seventy-five undergraduates from Kent State University participated for research credit in an introductory psychology class. Participants were randomly assigned to one of four delay groups (0, 1, 7, or 30), defined by the number of days between practice and test sessions.

Materials. Stimuli consisted of 120 AA problems (the maximum number of possible problems using the whole alphabet with addend sizes 2, 3, and 4, allowing for a few additional items to be used as examples during instructions). Half of the AA problems were false (e.g., A + 2 = D), and half were true (e.g., B + 2 = D). A third of the AA problems were of addend size 2 (e.g., A + 2 = C), a third were of addend size 3 (e.g., A + 3 = D), and a third were of addend size 4 (e.g., A + 4 = E).

Procedure. The AA problems were divided into five sets. Each set consisted of 12 true and 12 false problems, with equal numbers of true and false problems of each addend size. Sets were assigned to conditions (counterbalanced across participants) on the basis of the number of times items were repeated during practice. Of greatest importance for present purposes, two sets were assigned to the *low-repetition condition*, involving a relatively small number of repetitions per item (one set of items was repeated two times during practice and one set was repeated four times; analyses revealed no differences in performance for the two sets so they are collapsed into one condition for simplicity). Of lesser importance, two sets were assigned to the *high-repetition condition*, involving a larger number of repetitions per item (one set of items was repeated eight times during practice and one set was repeated 16 times; analyses revealed no differences in performance for the two sets so they are collapsed into one condition). The fifth set was assigned to the *novel-item condition*, in which each item was presented once during the test session.

The original intent for having the low- versus high-repetition conditions was to compare loss of item-general gains and item-specific gains within-subjects. In previous work (Grant, & Logan, 1993; Logan, & Klapp, 1991), 16 repetitions per item were sufficient for item-specific effects to emerge. To foreshadow, results for high-repetition items showed limited transfer to memory-based processing, but we report results for this condition nonetheless to further document effects of delays on practice gains.[2]

At the beginning of the practice session, participants read instructions and then completed six true and six false sample problems to become familiar with the task. The 12 examples were used only as warm-up problems and were not presented to participants during practice. Upon completion of the warm-up trials, participants were given feedback concerning their average response speed and accuracy. The experimenter then checked participants' performance to ensure they had completed the warm-up problems appropriately and to answer any questions about the procedure.

[2] The lack of transfer to item-specific processing during practice for the high-repetition items was surprising (given earlier work showing 16 trials were sufficient for this type of gain to occur). However, in previous research, participants practiced many fewer items than in the current research. Further exploration of possible set size effects may represent an interesting issue for future research.

Participants then continued with the experimental trials. During the practice session, participants completed 720 AA trials (the total from 24 problems repeated 2, 4, 8, or 16 times). The order of the 720 AA trials was randomized anew for each participant.

Participants clicked a button on the screen to start the practice session, which was then replaced by an orientation stimulus (***) presented for 500 ms. The orientation stimulus was then replaced by an AA problem and two response buttons (*TRUE* and *FALSE*) appearing below the AA problem. After participants clicked on a response button, the stimulus and the response buttons disappeared. If the response entered was incorrect, a red "ERROR" message was presented for 1,000 ms, followed by a button labeled *next*. If the participant's response was correct, the *next* button was immediately presented. The participant clicked on the *next* button to present the orientation stimulus for the next trial, followed by the next AA problem, and so on. Response time was recorded as the time between stimulus onset and clicking one of the two response buttons. Feedback of the same form as described for the warm-up trials was presented to participants after every 48 trials. Participants were also given the opportunity to take a small break during the presentation of feedback.

Upon completion of the practice session, participants in the 0-day delay group then completed the test session, with only a minimal delay while the experimenter reset the program on the computer. Participants in the remaining delay groups were dismissed and reminded to return at the specified interval.

At the beginning of the test session, all participants again read instructions and performed warm-up problems. Participants were then presented all 120 AA problems from the five sets in random order. Thus, the test session included one trial for each of the 96 previously practiced problems (48 low-repetition items and 48 high-repetition items) and one trial for each of 24 novel AA problems. The 120 AA problems were presented in the same manner as in the practice session, with the exception that participants did not receive performance feedback or opportunities to take a break.

Results

Data for nine participants were dropped from analyses because of failure to follow directions during the practice session, performance below 65% accuracy, failure to return for the test session, failure to exhibit any performance gains during practice, or evidence of practicing between the practice and test sessions. Accuracy was 91.5% (SE = 1.0%) during the practice session, and 91.9% (SE = 1.0%) and 92.6% (SE = 1.0%) for novel and practiced items during the test session, respectively. Analyses of response times were conducted on correct response trials only (excluding response times less than 50 ms and greater than 9,000 ms; <1% of trials). To minimize the effects of outliers, raw response times were first log-transformed and averaged over trials for each participant, and then individual averages were anti-log-transformed (cf. Rickard, 1997, 2007).

For each participant's data, response times in the low-repetition condition were averaged into miniblocks, with three consecutive trials averaged for the practice session data and two consecutive trials averaged for the test session data (to yield the same number of response times for both practiced items and novel items when plotting the test session data). Mean response times in each group for each miniblock in the practice and test sessions are presented in Figure 1. For the high-repetition condition, miniblocks were constructed with 12 consecutive trials averaged for the practice session (to yield the same number of miniblocks as in the low-repetition condition) and two consecutive trials averaged for the test session data. Mean response times in each group for each miniblock in the practice and test sessions are presented in Figure 2. In addition, for low- and high-repetition conditions, two-parameter power functions were fit to group antilog practice response times and extrapolated through trials in the test session to assess whether potentiation occurred.

Given that item-general gains are least likely to be contaminated with item-specific gains in the low-repetition condition, low-repetition results are of primary interest and will be reported first in each section. After low-repetition results, a summary of parallel high-repetition results will be reported.

To confirm that performance gains were obtained during practice, for the low-repetition items we compared mean response times across the first four miniblocks of practice with the last four miniblocks of practice (see the first two columns of Table 2). A 2 (miniblock: first four, last four) × 4 (delay group: 0 days, 1 day, 7 days, 30 days) mixed-factor analysis of variance (ANOVA) resulted in a significant main effect of miniblock, $F(1, 62) = 246.14$, $MSE = 307409$, $p < .001$. Response times were greater at the beginning of practice (M = 3,952 ms, SE = 112) than at the end of practice (M = 2,444 ms, SE = 92). Neither the main effect of group nor the Group × Miniblock interaction was significant (Fs < 1.71). For the high-repetition items, results of a 2 (miniblock: first four, last four) × 4 (delay group: 0 days, 1 day, 7 days, 30 days) mixed-factor ANOVA paralleled the results for the low-repetition items. The main effect of miniblock was significant, $F(1, 62) = 309.63$, $MSE = 340543$, $p < .001$. Response times were greater at the beginning of practice (M = 4,025 ms, SE = 114) than at the end of practice (M = 2,228 ms, SE = 73). Neither the main effect of group nor the Group × Miniblock interaction was significant (Fs < 2.09).

In the next two sections, we first report manipulation checks confirming that the observed gains were item-general practice gains. Second and most important, we then report the effects of delay on these item-general practice gains.

Manipulation checks. To confirm that the gains during practice were item-general, we first examined the effect of addend size on response times. The *addend slope* was calculated by fitting a linear slope across response times for items of addend size 2, 3, and 4. For each individual, we computed the addend slope for practiced items for the first trial of the practice session, for the last trial of the practice session, and for the test session. We also computed the addend slope for the novel items during the test session. Mean slopes across participants in each group are reported in Table 3. For the low-repetition practiced items, a 3 (trial: practice session first trial, practice session last trial, test session) × 4 (delay group: 0 days, 1 day, 7 days, 30 days) mixed-factor ANOVA revealed a significant main effect of trial, $F(2, 61) = 5.11$, $MSE = 54217$, $p = .009$, but neither the main effect of delay group nor the interaction was significant. Accordingly, we collapse across delay group in subsequent analyses.

The addend slope at the end of practice was still significantly greater than zero (M = 342 ms/addend, SE = 31), $t(65) = 11.15$, $p < .001$. As outlined in the introduction, a slope greater than zero

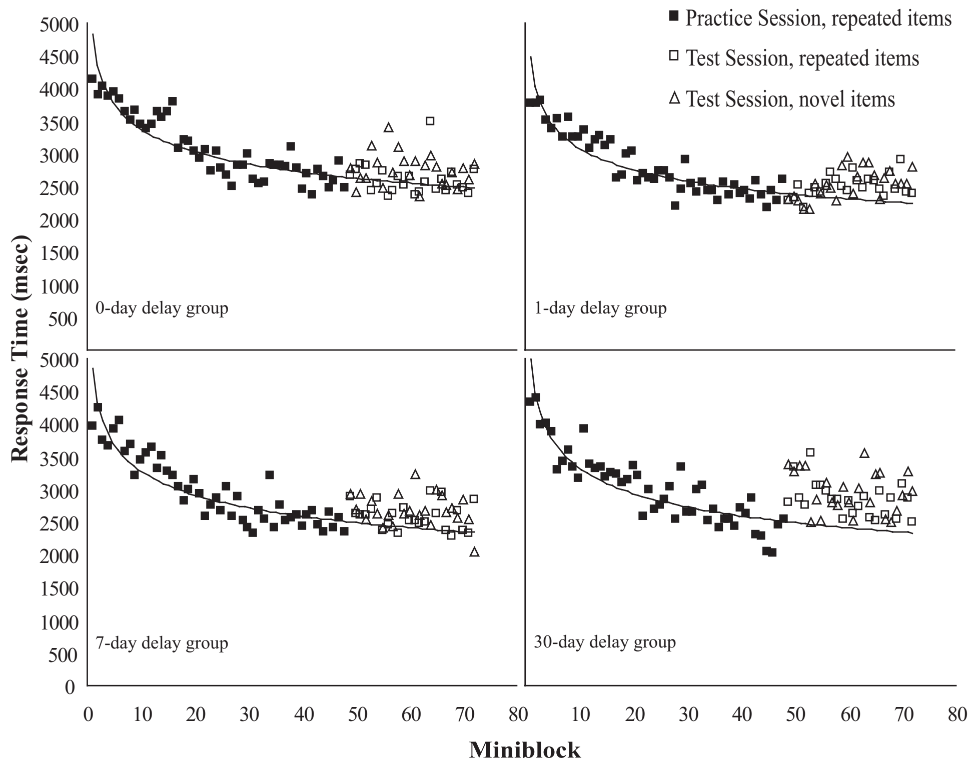

Figure 1. Experiment 1 mean response times as a function of miniblocks of trials and experimental session for each delay group for low-repetition items. Test session response times are further separated into novel-item and repeated-item miniblocks. Lines represent the best-fitting power function across group mean response times in the practice session, extrapolated across trials in the test session.

at the end of practice is consistent with the idea that items are still being processed via algorithm. However, the addend slope at the end of practice was significantly lower than at the beginning of practice ($M = 475$ ms/addend, $SE = 30$), $t(65) = 3.09$, $p = .003$. On the one hand, the reduction in slope is consistent with the idea that the efficiency of the item-general algorithm used to compute responses improved with practice. On the other hand, a reduction in slope may also indicate that partial transfer from item-general processing to item-specific processing occurred. However, the finding that the addend slope did not significantly differ between novel items ($M = 368$ ms/addend, $SE = 26$) and practiced items at the end of practice or during test ($M = 384$ ms/addend, $SE = 27$; ts < 0.66) strongly suggests that reduction in slope was due to increased efficiency of the item-general algorithm, rather than partial transfer to item-specific processing. Finally, the addend slope for practiced items was not significantly different between the end of practice and the test session, $t(65) = 1.00$, $p = .32$.

For the high-repetition items, a 3 (trial: practice session first trial, practice session last trial, test session) × 4 (delay group: 0 days, 1 day, 7 days, 30 days) mixed-factor ANOVA revealed a significant main effect of trial and a Trial × Delay Group interaction that approached significance, $F(2, 124) = 14.30$, $MSE = 46482$, $p < .001$, and $F(6, 124) = 2.09$, $MSE = 46482$, $p = .059$, respectively. Results of follow-up tests largely paralleled results for the low-repetition items, except that the addend slope for practiced items significantly differed between the end of practice ($M = 320$ ms, $SE = 31$) and the test session ($M = 398$ ms, $SE = 26$). An increase in slope across a delay is consistent with the idea that during practice, a partial transfer to item-specific processing may have occurred, with some loss across the delay. However, the result is also consistent with the occurrence of loss of algorithm efficiency across the delay.

To show converging evidence that gains during practice were item-general, we also compared mean response times between novel items and practiced items during test (see Table 4). For the low-repetition condition, response times were significantly different between novel and practiced items for the 0-day delay group, $t(18) = 2.58$, $p = .019$. However, for all other groups, no significant differences were found (all ts < 1.38, all ps $> .189$). For the high-repetition condition, response times were significantly greater for novel than for practiced items for all delay groups (all ts $>$ 2.35, all ps $< .033$), which adds further evidence for the possibility that partial transfer to item-specific processing occurred during practice for the high-repetition items.

Having established that gains made during practice were largely item-general for the low-repetition items, we turn now to the question of greatest interest: What are the effects of a delay on item-general practice gains?

Effects of delay on practice gains. To ascertain the effects of a delay on item-general practice gains, we first computed PGL values based on group means for comparison with group PGL estimates from prior studies summarized in Table 1. For the

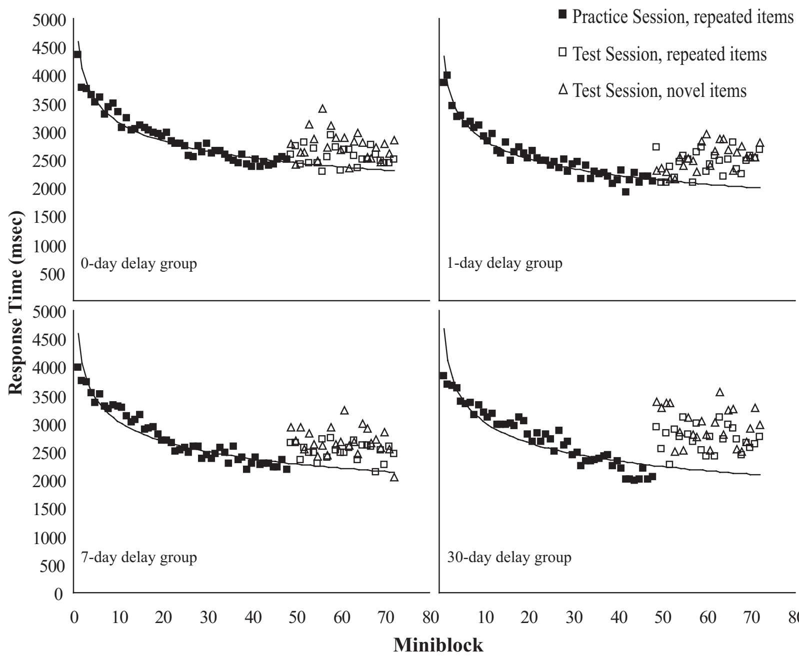

Figure 2. Experiment 1 mean response times as a function of miniblocks of trials and experimental session for each delay group for high-repetition items. Test session response times are further separated into novel-item and repeated-item miniblocks. Lines represent the best-fitting power function across group mean response times in the practice session, extrapolated across trials in the test session.

low-repetition condition, gain was calculated by subtracting the mean for the last four miniblocks of practice from the mean for the first four miniblocks of practice, loss was calculated by subtracting the mean for the last four miniblocks of practice from the mean for the first six miniblocks during the test session, and PGL was calculated as (loss/gain) × 100. Group-level PGL values were 6.2, −0.9, 16.5, and 43.4% for the 0-, 1-, 7-, and 30-day delay groups, respectively. For the high-repetition condition, group-level PGL values were calculated on the basis of the first and last miniblocks (so that estimates are based on 12 trials as in the low-repetition calculations) and were 5.3, 11.8, 18.9, and 35.2% for the 0-, 1-, 7-, and 30-day delay groups, respectively.

Unlike in the previous studies mentioned earlier, we were also able to compute PGL values at the level of the individual. Mean gain, loss, and PGL across individuals in each of the delay groups are reported in Table 2. For the low-repetition condition, only the mean PGL value for the 30-day delay group was significantly different from zero, $t(14) = 3.19$, $p = .007$; for all others, $t\text{s} < 1.04$. For the high-repetition condition, the mean PGL value for the 7-day delay group was significantly greater than zero, $t(15) = 3.07$, $p = .008$, and the PGL for the 30-day delay group was marginally greater than zero, $t(14) = 1.87$, $p = .082$.

Compared with the PGL values derived from previous studies, PGL values here were surprisingly low, suggesting that item-

Table 2
Summary of Response Time Statistics for Practiced Items in Experiment 1

Item type and delay	Practice start	Practice end	Test start	Gain	Loss	PGL
Low repetition						
0 days	4,011 (217)	2,600 (157)	2,688 (143)	1,411 (174)	88 (82)	0.5 (5.3)
1 day	3,721 (232)	2,372 (170)	2,360 (200)	1,349 (172)	−11.8 (181)	−10.3 (9.9)
7 days	3,904 (214)	2,488 (132)	2,721 (161)	1,416 (189)	234 (157)	6.5 (13.3)
30 days	4,177 (242)	2,267 (276)	3,095 (216)	1,910 (237)	828 (252)	31.1 (9.8)
High repetition						
0 days	4,358 (253)	2,504 (116)	2,601 (137)	1,854 (179)	98 (115)	3.9 (9.2)
1 day	3,855 (235)	2,108 (138)	2,315 (152)	1,747 (200)	207 (113)	11.3 (8.1)
7 days	3,979 (155)	2,182 (109)	2,522 (129)	1,797 (129)	340 (106)	19.1 (6.2)
30 days	3,834 (238)	2,055 (200)	2,682 (147)	1,779 (290)	627 (226)	21.2 (11.3)

Note. All except PGL data are measured in milliseconds. Practice start, practice end, and test start are means across individuals in each group. Gain is the means across individual differences between practice start and practice end in each group. Loss is the means across individual differences between test start and practice end in each group. PGL is computed as (loss/gain) × 100 for each individual, and values represent means across individuals in each group. Standard errors are reported in parentheses. PGL = percentage of gain lost.

Table 3
Addend Slopes in Experiment 1

Item type and delay	Practiced items First	Practiced items Last	Practiced items Test	Novel items test
Low repetition				
0 days	457 (62)	343 (39)	411 (51)	335 (55)
1 day	578 (44)	305 (84)	404 (49)	367 (49)
7 days	565 (60)	451 (59)	314 (65)	414 (55)
30 days	401 (65)	266 (58)	403 (52)	360 (48)
High repetition				
0 days	429 (50)	382 (34)	388 (53)	335 (55)
1 day	409 (82)	227 (68)	383 (40)	367 (49)
7 days	573 (77)	390 (58)	397 (49)	414 (55)
30 days	654 (62)	265 (87)	425 (69)	360 (48)

Note. Values are slopes indicating mean millisecond increase in response time per unit increase in addend size. Standard errors are in parentheses.

general practice gains are relatively resistant to loss across a delay. For example, the degree of loss observed after 30 days here was apparent after only a one-day delay in Anderson et al.'s (1999) Experiment 2 (see Table 1).

Of secondary interest, we also examined potentiation after a delay (Rickard, 2007). Inspection of Figures 1 and 2 indicates that potentiation did not occur for item-general practice gains across a delay. Response times during the test session did not consistently fall below the extrapolated fit from response times in the practice session in any group. These results provide further evidence that potentiation may not be robust (a point we revisit in the General Discussion section).

Experiment 2

Compared with the loss of item-specific gains in previous research (see Table 1), relatively good preservation of item-general practice gains across a delay were seen in Experiment 1. However, comparison of the loss of item-general practice gains from Experiment 1 with the loss of item-specific gains in other studies must be made with caution. One limitation of such a comparison concerns the wide variety of tasks used across the different studies, given that these tasks most likely differ along other dimensions, such as difficulty.

Accordingly, the goal of Experiment 2 was to more directly compare the loss of item-general practice gains with the loss of item-specific practice gains. Experiment 2 included two groups: an item-general practice group and an item-specific practice group. Both groups performed the same AA verification task used in Experiment 1. However, the number of specific AA problems presented to participants differed between the two groups. Participants in the item-general practice group practiced a large number of unique AA problems that were each repeated a small number of times. Participants in the item-specific practice group practiced a small number of unique AA problems that were each repeated many times. For each group, the total number of AA trials was the same, and the delay between the practice session and the test session was 2 days. We predicted, on the basis of the outcomes of Experiment 1, significantly lower PGL in the item-general practice group than in the item-specific practice group.

Method

Participants and design. Sixty-two undergraduates from Kent State University participated for research credit in an introductory psychology class. Participants were randomly assigned to one of two practice groups (item-general practice or item-specific practice).

Materials and procedure. Experimental stimuli were the same 120 AA items used in Experiment 1. Stimuli were divided into 20 sets of six items. Each set consisted of three true items and three false items, one of each type at each addend size (2, 3, and 4). Assignment of item sets to practice and test conditions (described next) was counterbalanced across participants.

During the practice session, participants in the item-general practice group were presented with 96 unique AA items, presented once in each of three blocks of practice (288 trials). Presentation order of the 96 items was randomized anew for each block. Participants in the item-specific practice group were presented with six unique AA items, presented once in each of 48 blocks of practice (288 trials). Presentation order was randomized anew for each block.

All participants returned 2 days later for the test session, in which they completed 120 trials. The item-general practice group was presented with the 96 previously practiced items and with 24 novel items (one trial each, in random order). The item-specific practice group was presented with the six previously practiced items (16 times each) and with 24 novel items (one trial each, in random order). All other aspects of the procedure were identical to those in Experiment 1.

Results

Data for 11 participants were dropped from analyses because of failure to complete both sessions of the experiment, performance below 65% accuracy, receipt of wrong test material, failure to show evidence of learning during the practice session, or evidence of practicing between the practice and test sessions. Accuracy was 95.4% ($SE = 0.7$) during the practice session and 92.9% ($SE = 1.0$) and 96.5% ($SE = 0.6$) for novel and practiced items during the test session, respectively. Analyses of response times were conducted on correct response trials only (excluding response times less than 50 ms and greater than 9,000 ms; $<1\%$), using the log/antilog transformation procedure described in Experiment 1.

For each participant's data, response times for practiced items were averaged into miniblocks, with six consecutive trials averaged for the practice session data and four consecutive trials

Table 4
Response Times in Test Session in Experiment 1

Delay	Practiced items Low repetition	Practiced items High repetition	Novel items
0 days	2,614 (124)	2,546 (123)	2,756 (120)
1 day	2,505 (182)	2,436 (161)	2,520 (155)
7 days	2,595 (130)	2,502 (123)	2,677 (147)
30 days	2,836 (156)	2,715 (156)	2,931 (167)

Note. Values are means in milliseconds. Standard errors are in parentheses.

averaged for the test session data (to yield the same number of miniblocks for both practiced items and novel items when plotting the test session data). Mean response times in each group for each miniblock in the practice and test sessions are presented in Figure 3. In addition, for each group we have plotted the two-parameter power function fit to group antilog practice response times and extrapolated through trials in the test session to assess potentiation.

To confirm performance gains during practice, we compared mean response times across the first two miniblocks during practice to mean response times across the last two miniblocks during practice (see first two columns of Table 5). A 2 (miniblock: first two, last two) × 2 (practice group: item-specific, item-general) mixed-factor ANOVA resulted in significant main effects of miniblock and practice group, $F(1, 49) = 242.95$, $MSE = 473820$, $p < .001$, and $F(1, 49) = 40.71$, $MSE = 600121$, $p < .001$, respectively. The interaction was also significant, $F(1, 49) = 18.41$, $MSE = 473820$, $p < .001$, indicating that practice gains were considerably greater for the item-specific practice group than for the item-general practice group.

In the next two sections, we first report manipulation checks confirming item-general versus item-specific practice gains in the two groups. Second and most important, we then report the effects of the delay on item-general practice gains and compare these with effects on item-specific gains.

Manipulation checks. To confirm that practice gains were item-general for the item-general practice group and item-specific for the item-specific practice group, we first examined the effect of addend size on response times. The addend slope was computed as in Experiment 1. All slopes are reported in Table 6. For the practiced items, a 3 (trial: practice session first trial, practice session last trial, test session) × 2 (practice group: item-general, item-specific) mixed-factor ANOVA revealed a significant main effect of trial, $F(2, 48) = 10.61$, $MSE = 82250$, $p < .001$; a significant main effect of practice group, $F(1, 49) = 27.89$, $MSE = 97824$, $p < .001$; and a significant interaction between trial and practice group, $F(2, 98) = 6.76$, $MSE = 82250$, $p = .002$. Accordingly, we separate by practice group in the analyses that follow.

For the item-general practice group, the addend slope at the end of practice was still significantly greater than zero (M = 410 ms/addend, SE = 47), $t(24) = 8.81$, $p < .001$. To revisit, a slope greater than zero at the end of practice is consistent with the idea that items are still being processed via algorithm. However, the addend slope declined from the beginning of practice (M = 524 ms/addend, SE = 93) to the end of practice, although not significantly, $t(24) = 1.14$, $p = .27$. On the one hand, the reduction in slope is consistent with the idea that the efficiency of the item-general algorithm used to compute responses improved with practice. On the other hand, a reduction in slope may also indicate that partial transfer from item-general processing to item-specific processing occurred. However, the finding that the addend slope did not significantly differ between novel items (M = 414 ms/addend, SE = 48) and practiced items at the end of practice or during test (M = 407 ms/addend, SE = 30; $ts < 0.11$) strongly suggests that reduction in slope was due to increased efficiency of the item-general algorithm, rather than partial transfer to item-specific processing. Finally, the addend slope for practiced items was not significantly different between the end of practice and the test session, $t(24) = 0.05$, $p = .96$.

In contrast, for the item-specific practice group, the addend slope at the end of practice was not significantly different from zero (M = 1 ms/addend, SE = 27), $t(25) = 0.03$, $p = .97$, consistent with the idea that items were being processed via memory-based processing. The addend slope at the end of practice was significantly lower than at the beginning of practice (M = 496 ms/addend, SE = 88), $t(25) = 5.71$, $p < .001$, also consistent with transfer to item-specific processing during practice. As further evidence, the addend slope significantly differed between novel items (M = 477 ms/addend, SE = 87) and practiced items at the end of practice and during test (M = 43 ms/addend, SE = 19), $t(25) = 4.92$, $p < .001$, and $t(25) = 5.20$, $p < .001$, respectively. Finally, the addend slope for practiced items was not significantly different from the end of practice to the test session, $t(25) = 0.11$, $p = .19$.

To show converging evidence for the type of gain in each group, we also compared mean response times between novel items and practiced items during test (see Table 7). In the item-general practice group, response times were not significantly different between novel and practiced items, $t(24) = 1.63$, $p = .12$. In the item-specific practice group, response times were significantly greater for novel versus practiced items, $t(25) = 14.51$, $p < .001$.

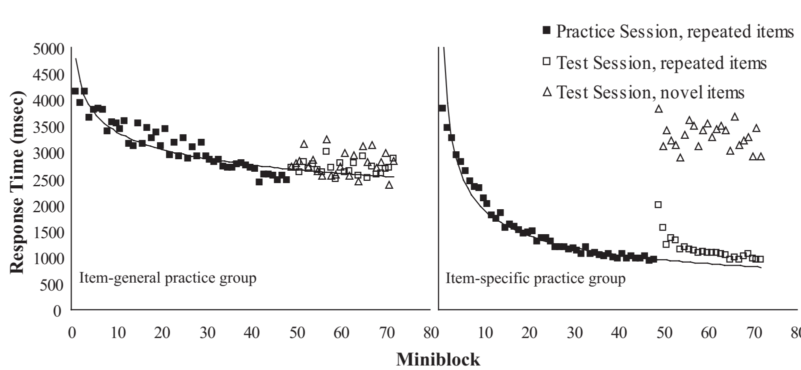

Figure 3. Experiment 2 mean response times as a function of miniblocks of trials and experimental session for each practice group. Test session response times are further separated into novel-item and repeated-item miniblocks. Lines represent the best-fitting power function across group mean response times in the practice session, extrapolated across trials in the test session.

Table 5
Summary of Response Time Statistics for Practiced Items in Experiment 2

Practice group	Practice start	Practice end	Test start	Gain	Loss	PGL
Item-general	4,046 (212)	2,506 (124)	2,677 (105)	1,540 (222)	91 (79)	6.7 (7.1)
Item-specific	3,652 (155)	942 (38)	1,599 (92)	2,710 (161)	657 (95)	25.9 (3.8)

Note. All except PGL data are measured in milliseconds. Practice start, practice end, and test start are means across individuals in each group. Gain is the means across individual differences between practice start and practice end in each group. Loss is the means across individual differences between test start and practice end in each group. PGL is computed as (loss/gain) × 100 for each individual, and values represent means across individuals in each group. Standard errors are reported in parentheses. PGL = percentage of gain lost.

Having established that gains made during practice were item-general for the item-general practice group and item-specific for the item-specific practice group, we turn now to the questions of greatest interest: What are the effects of a delay on item-general practice gains, and how do they compare with the effects of delay on item-specific practice gains?

Effects of delay on practice gains. We first computed PGL values based on group means for comparison with group PGL estimates from prior research. Gain was calculated by subtracting the mean for the last two miniblocks of practice from the mean for the first two miniblocks of practice, loss was calculated by subtracting the antilog mean for the last two miniblocks of practice from the first three miniblocks during the test session, and PGL was calculated as (loss/gain) × 100. Group-level PGL values were 5.9% and 24.2% for the item-general and item-specific practice groups, respectively.

As in Experiment 1, we also computed PGL values for each individual (see Table 5). Only the mean PGL value for the item-specific practice group was significantly greater than zero, $t(25) = 6.82$, $p < .001$. The PGL in the item-general practice group after a 2-day delay ($M = 6.7\%$, $SE = 7.1$) was within the range that would be expected on the basis of the outcomes of Experiment 1. In contrast, the PGL in the item-specific practice group ($M = 25.9\%$, $SE = 3.8$) was sizeable after a 2-day delay, consistent with reported loss of item-specific gains across 1- to 2-day delays in previous research (32%–41%). Together, the PGL values suggest that loss of practice gains across a delay is more modest for item-general gains than for item-specific gains. Indeed, the PGL values were significantly different between the practice groups, $t(49) = 2.42$, $p = .019$.

We also examined, of secondary interest, potentiation after a delay. Inspection of Figure 3 indicates that potentiation did not occur in either group. Response times during the test session did not consistently fall below the extrapolated fit from response times in the practice session in either group. These results provide further evidence that potentiation may not be robust.

Table 6
Addend Slopes in Experiment 2

Practice group	Practiced items: First	Practiced items: Last	Practiced items: Test	Novel items test
Item-general	524 (93)	410 (47)	407 (30)	414 (48)
Item-specific	496 (88)	1 (27)	43 (19)	477 (87)

Note. Values are slopes indicating mean millisecond increase in response time per unit increase in addend size. Standard errors are in parentheses.

Experiment 3

The results of Experiment 2 support the conclusion that item-general gains are less susceptible to loss across a delay than are item-specific gains. However, an alternative account is that item-general and item-specific gains do not differ in loss across a delay but instead are differentially affected by switch costs during the test session. *Switch cost* refers to a response time increase for a given trial when the task performed on the previous trial is different versus when the task performed on the previous trial is the same (e.g., Koch & Allport, 2006; Logan & Bundesen, 2003; Meiran, 1996; Rogers & Monsell, 1995; Yeung, 2010). In the item-general practice group in Experiment 2, practiced and novel items were both solved via algorithm during the test session. In contrast, participants in the item-specific practice group were presumably switching between memory retrieval for practiced items and algorithmic processing for novel items during the test session. Given that switch costs would occur only for the item-specific practice group, the observed pattern may reflect a switch cost decrement occurring during test for the item-specific practice group but not for the item-general practice group rather than differential loss across a delay.

To evaluate the extent to which switch costs may have differentially biased estimates of the PGL for the item-general and item-specific practice groups, in Experiment 3 we did not present half of the participants in each group with novel items during the test session. If switch costs influenced the PGL for the item-specific practice group, we would expect a significant interaction between practice group (item-general vs. item-specific) and novel-item group (present vs. absent), with greater PGL differences between novel items present versus absent in the item-specific practice group than in the item-general group. Alternatively, if switch costs do not bias PGL values for either practice group, we

Table 7
Response Times in Test Session in Experiment 2

Practice group	Practiced items	Novel items
Item-general	2,455 (96)	2,313 (114)
Item-specific	1,096 (32)	2,898 (116)

Note. Values are group means in milliseconds. Standard errors are in parentheses.

would expect neither a main effect of novel-item group nor an interaction of novel-item group with practice group.

Experiment 3 included one additional modification to the Experiment 2 method. In the item-specific group in Experiment 2, the number of other items intervening between repetitions of a given item could have been as few as zero and on average about five. Therefore, on some trials, items may still have been in working memory rather than retrieved from long-term memory. If so, response times would have been faster, which would have led to overestimating speed of retrieval from long-term memory during practice. Overestimating speed of retrieval during practice would inflate estimates of gain during practice and loss across the delay. To minimize the possibility of retrieval from working memory in Experiment 3, we ensured that a minimum of six items intervened between repetitions during practice.

Method

Participants and design. Ninety-nine undergraduates from Kent State University participated for research credit in an introductory psychology class. Participants were randomly assigned to four groups, based on a 2 (practice group: item-general, item-specific) $\times$ 2 (novel-item group: present, absent) factorial design.

Materials and procedure. Experimental stimuli were the same 120 AA items as in Experiments 1 and 2. Stimuli were divided into 20 sets of six items. Each set consisted of three true items and three false items, one of each type at each addend size (2, 3, and 4). Assignment of item set to the practice and test conditions (described next) was counterbalanced across participants.

During the practice session, participants in the item-general practice group were presented with 96 AA items, presented once in each of five blocks of practice (480 trials). Presentation order of the 96 items was randomized anew for each block. Participants in the item-specific practice group were presented with two sets of six AA items (12 items in total), presented once in each of 40 blocks of practice (480 trials). Presentation of items within each set was randomized anew for each block, but the presentation order of the two sets was the same in each practice block (Set 1 followed by Set 2) to ensure a minimum of six intervening items between trials for a given item.

All participants returned 2 days later for the test session. For the item-general practice group, participants were presented with the 96 previously practiced items for one trial each in random order. For the item-specific group, participants were presented with the 12 previously practiced items for eight trials each in random order. For half of the participants in each group, 24 novel items were randomly presented with the previously practiced items. All other aspects of the procedure were identical to those in Experiment 2.

Results

Data for eight participants were dropped from analyses because of failure to complete both sessions of the experiment, performance below 65% accuracy, failure to show evidence of learning during the practice session, or evidence of practicing between the practice and test sessions. Accuracy was 95.0% (SE = 0.5) during the practice session and 94.0% (SE = 1.0) and 96.7% (SE = 0.4) for novel and practiced items during the test session, respectively. Response times were trimmed ($<$1%) and transformed as in Experiments 1 and 2.

For each participant's data, response times for practiced items were averaged into miniblocks, with 12 consecutive trials averaged for the practice session data and four consecutive trials averaged for the test session data. Mean response times in each group for each miniblock in the practice and test sessions are presented in Figure 4. In addition, for each group we plotted the two-parameter

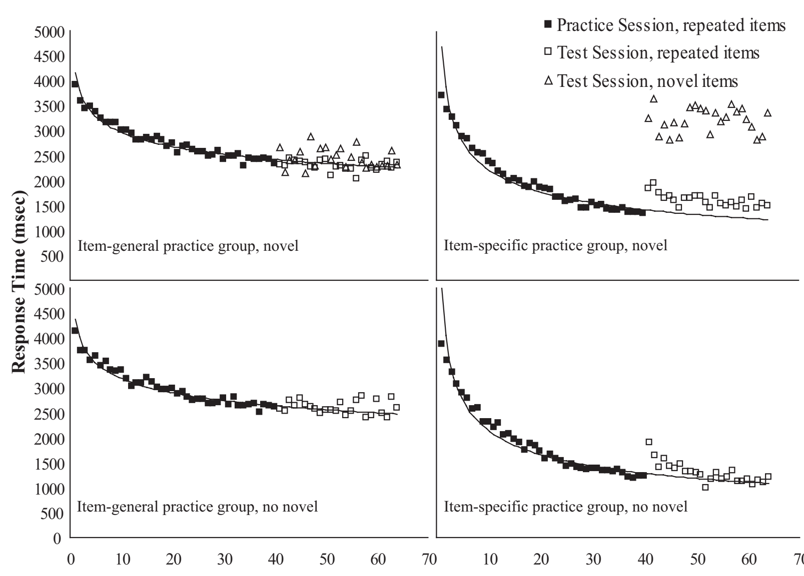

Figure 4. Experiment 3 mean response times as a function of miniblocks of trials and experimental session for each practice group. Test session response times are further separated into novel-item and repeated-item miniblocks. Lines represent the best-fitting power function across group mean response times in the practice session, extrapolated across trials in the test session.

power function fit to group response times for the practice session and extrapolated through trials in the test session to assess potentiation.

To confirm performance gains during practice, we compared mean response times across the first two miniblocks during practice with mean response times across the last two miniblocks during practice (see first two columns of Table 8). A 2 (miniblock: first two, last two) × 2 (practice group: item-specific, item-general) × 2 (novel-item group: present, absent) mixed-factor ANOVA resulted in significant main effects of miniblock and practice group, $F(1, 87) = 480.70$, $MSE = 387170$, $p < .001$, and $F(1, 87) = 25.06$, $MSE = 892370$, $p < .001$, respectively. The Miniblock × Practice Group interaction was also significant, $F(1, 87) = 27.54$, $MSE = 387170$, $p < .001$. Practice gains were considerably greater for the item-specific practice group than for the item-general practice group. Neither the main effect nor any interactions with novel-item group were significant (all Fs < 1.81).

In the next two sections, we first report manipulation checks confirming item-general versus item-specific practice gains in the two practice groups. Second and most important, we then report the effects of delay on item-general practice gains and compare these with effects on item-specific gains. Third, we evaluate the effect of presenting novel items during test to assess the extent to which switch costs may have influenced the primary outcomes.

Manipulation checks. To confirm that practice gains were item-general for the item-general practice group and item-specific for the item-specific practice group, we first examined the effect of addend size on response times. The addend slope was computed as in Experiment 1. The mean slopes are reported in Table 9. For the practiced items, a 3 (trial: practice session first trial, practice session last trial, test session) × 2 (practice group: item-general, item-specific) × 2 (novel-item group: present, absent) mixed-factor ANOVA revealed a significant main effect of trial, $F(2, 174) = 11.87$, $MSE = 125796$, $p < .001$; a significant main effect of practice group, $F(1, 87) = 15.28$, $MSE = 130279$, $p < .001$; and a significant interaction between trial and practice group, $F(2, 174) = 8.73$, $MSE = 125796$, $p < .001$. Neither the main effect nor any interactions with novel-item group were significant (all Fs < 0.85). Accordingly, we separate by practice group in the analyses that follow and collapse across novel-item group when applicable.

For the item-general practice group, the addend slope at the end of practice was still significantly greater than zero ($M = 349$ ms/addend, $SE = 56$), $t(46) = 6.21$, $p < .001$, consistent with the idea that items are still being processed via algorithm. In addition, the addend slope declined only slightly from the beginning of practice ($M = 385$ ms/addend, $SE = 69$) to the end of practice, $t(24) = 0.43$, $p = .67$. For participants with novel items present during test, the addend slope did not significantly differ between novel items ($M = 307$ ms/addend, $SE = 41$) and practiced items ($M = 317$ ms/addend, $SE = 55$) at the end of practice or during test ($M = 347$ ms/addend, $SE = 34$; ts < 1.30). Finally, for practiced items, the addend slope was not significantly different between the end of practice and the test session, $t(46) = 0.39$, $p = .70$.

In contrast, for the item-specific practice group, the addend slope at the end of practice was not significantly different from zero ($M = 5$ ms/addend, $SE = 37$), $t(43) = 0.13$, $p = .90$, consistent with the idea of transfer to item-specific memory-based processing. The addend slope at the end of practice was significantly lower than at the beginning of practice ($M = 463$ ms/addend, $SE = 78$), $t(43) = 4.87$, $p < .001$. For participants who received novel items during test, the addend slope significantly differed between novel items ($M = 253$ ms/addend, $SE = 99$) and practiced items at the end of practice ($M = -46$ ms/addend, $SE = 41$), $t(21) = 3.36$, $p = .003$, but not between novel items and practiced items during test ($M = 131$ ms/addend, $SE = 34$), $t(21) = 1.13$, $p = .272$. Taken together, these two results suggest that practiced items were processed via item-specific memory retrieval during practice, but then partial reversion back to algorithm-based processing occurred for some practiced items during test (which would be consistent with forgetting across a delay). In support of this

Table 9
Addend Slopes in Experiment 3

Practice group and novel	Practiced items			Novel items test
	First	Last	Test	
Item-general				
Present	422 (97)	317 (55)	347 (34)	307 (41)
Absent	351 (100)	379 (97)	389 (33)	
Item-specific				
Present	483 (104)	–46 (40)	131 (34)	253 (99)
Absent	443 (118)	55 (60)	112 (38)	

Note. Values are slopes indicating mean millisecond increase in response time per unit increase in addend size. Standard errors are in parentheses.

Table 8
Summary of Response Time Statistics for Practiced Items in Experiment 3

Practice group and novel	Practice start	Practice end	Test start	Gain	Loss	PGL
Item-general						
Present	3,908 (196)	2,343 (103)	2,349 (135)	1,564 (147)	6 (69)	–3.7 (4.6)
Absent	4,125 (224)	2,611 (122)	2,635 (168)	1,514 (144)	24 (106)	2.1 (9.9)
Item-specific						
Present	3,726 (182)	1,347 (92)	1,852 (112)	2,379 (201)	506 (96)	16.9 (4.4)
Absent	3,873 (242)	1,236 (94)	1,651 (85)	2,637 (238)	416 (77)	18.3 (3.8)

Note. All except PGL data are measured in milliseconds. Practice start, practice end, and test start are means across individuals in each group. Gain is the means across individual differences between practice start and practice end in each group. Loss is the means across individual differences between test start and practice end in each group. PGL is computed as (loss/gain) × 100 for each individual, and values represent means across individuals in each group. Standard errors are reported in parentheses. PGL = percentage of gain lost.

idea, the addend slope for practiced items was significantly lower at the end of practice ($M = 5$ ms/addend, $SE = 37$) than in the test session ($M = 121$ ms/addend, $SE = 25$), $t(43) = 3.76$, $p = .001$.

To show converging evidence for the type of gain in each practice group, we also compared mean response times between novel items and practiced items during test (see Table 10). In the item-general practice group, response times did not significantly differ between novel and practiced items, $t(22) = 1.50$, $p = .21$. In the item-specific practice group, response times were significantly greater for novel versus practiced items, $t(21) = 8.68$, $p < .001$.

Having established that gains made during practice were item-general for the item-general practice group and item-specific for the item-specific practice group, we turn now to two important questions: What are the effects of a delay on item-general practice gains, and how do they compare with the effects of delay on item-specific practice gains?

Effects of delay on practice gains. For comparison with group PGL estimates from prior research, group-level PGL values are reported first. Gain was calculated by subtracting the mean for the last miniblock of practice from the mean for the first miniblock of practice, loss was calculated by subtracting the mean for the last miniblock of practice from the first three miniblocks during the test session, and PGL was calculated as (loss/gain) × 100. Group-level PGL values for the item-general practice group were –0.4% and 1.6%, respectively, for novel items present and novel items absent. In contrast, group-level PGL values for the item-specific practice group were 21.3% and 15.8%, respectively, for novel items present and novel items absent.

As in Experiments 1 and 2, we also computed PGL values for each individual (see Table 8). Only the mean PGL values for the item-specific practice group were significantly greater than zero, $t(21) = 3.88$, $p = .001$, and $t(21) = 4.86$, $p < .001$ for the novel present and novel absent conditions, respectively. A 2 (practice group: item-general, item-specific) × 2 (novel-item group: present, absent) factorial ANOVA revealed a significant main effect of practice group, $F(1, 87) = 8.38$, $MSE = 922$, $p = .005$, which suggests that loss of practice gains across a delay is more modest for item-general gains ($M = –0.7\%$, $SE = 5.5$) than for item-specific gains ($M = 17.6\%$, $SE = 2.9$). Neither the main effect of novel-item group nor the interaction was significant (all Fs < 0.33), which suggests that PGL values were not influenced by the presence of novel items for either type of practice group.

Although these results suggest that novel items do not influence PGL values, examination of Figure 4 reveals that novel items may have affected overall response times for practiced items during the test session. To examine the effect of novel items on practiced item performance, we performed a 2 (practice group: item-general, item-specific) × 2 (novel-item group: present, absent) factorial ANOVA on overall mean response time for practiced items during test. Results revealed a significant main effect of practice group, $F(1, 87) = 82.22$, $MSE = 271480$, $p < .001$, indicating that response times for practiced items were slower for the item-general practice group ($M = 2{,}391$ ms, $SE = 90$) than for the item-specific practice group ($M = 1{,}397$ ms, $SE = 65$). Although the main effect of novel-item group was not significant, the interaction was significant, $F(1, 87) = 7.19$, $MSE = 271480$, $p = .009$. For the item-general practice group, response times for practiced items during test were not significantly different when novel items were present ($M = 2{,}254$ ms, $SE = 118$) versus absent ($M = 2{,}522$ ms, $SE = 132$), $t(45) = 1.50$, $p = .14$. For the item-specific practice group, response times for practiced items during test were significantly greater when novel items were present ($M = 1{,}556$ ms, $SE = 98$) versus absent ($M = 1{,}237$ ms, $SE = 72$), $t(42) = 2.62$, $p = .012$, consistent with the occurrence of switch costs. Why did novel items influence overall response time but not PGL in the item-specific group? The likely explanation for why PGL was not biased by novel items concerns the proportion of trials that included novel items at the beginning of test. Given that only 24 of a total of 120 trials during test consisted of novel items, most participants experienced very few (if any) novel trials during the first 12 trials of the test session ($M = 2.2\%$), which were the basis for computing PGL.

Of secondary interest, we also examined potentiation after a delay. Inspection of Figure 4 indicates that potentiation did not occur in either practice group. However, in the item-specific practice group, response times for practiced items during test were closer to the extrapolated fit line by the end of test when novel items were absent. This result suggests that excluding novel items may be an important condition for obtaining evidence of potentiation in future research.

Table 10
Response Times in Test Session in Experiment 3

Practice group	Practiced items: Novel absent	Practiced items: Novel present	Novel items
Item-general	2,522 (132)	2,254 (118)	2,303 (119)
Item-specific	1,237 (72)	1,556 (98)	3,070 (152)

Note. Values are group means in milliseconds. Standard errors are in parentheses.

General Discussion

The present research represents the first direct examination of the effects of a delay on item-general practice gains in a cognitive task. Although some loss was observed, the key findings revealed that item-general practice gains are relatively robust across delays, particularly compared with the more severe effects of delay on item-specific practice gains (as observed in prior research and in Experiments 2 and 3).

Although no prior research on cognitive skill acquisition has directly examined the loss of item-general gains and how it compares with the loss of item-specific practice gains, to what extent might the current findings parallel results from basic memory research? For example, fuzzy-trace theory (Brainerd, Reyna, & Kneer, 1995; Reyna & Kiernan, 1994) states that encoding a stimulus event into memory involves the creation of two separate memory traces: a gist trace and a verbatim trace. Gist traces represent the meaning of the stimulus event, whereas verbatim traces represent the exact form of the stimulus event's surface features. When a stimulus is next encountered, both types of memory traces can be retrieved. Of potential relevance here, Reyna and Kiernan's (1994) results suggest that loss across a delay is greater for verbatim traces than for gist traces, which suggests that after a sufficient delay, traces available for retrieval would be mostly gist traces. Although the results of greater loss of verbatim traces than gist traces across a delay seems to parallel our finding of greater item-specific loss than item-general loss across a delay, only one of our processing routes—the item-specific processing route—is based on item-specific memory traces. However, the greater rate of loss for verbatim traces than for gist traces may explain the rapid loss of item-specific gains found in our studies. A straightforward extension of fuzzy-trace theory would assume that both verbatim and gist traces participate in memory-based processing of repeated items during practice, whereas processing of repeated items during test would be based primarily on just gist traces. Consistent with memory-based theories of automaticity, fewer traces would lead to slower response times.

Another finding from the verbal learning literature that might appear to parallel the current results involves the spacing effect, in which shorter versus longer lags between practice trials produce faster learning but greater loss (for review, see Cepeda, Pashler, Vul, Wixted, & Rohrer, 2006). Consistent with this finding, in Experiments 2 and 3, the lag between repetitions during practice was shorter for the item-specific group than for the item-general practice group at the level of a given item. However, for practice involving an item-general algorithm, defining lag at the level of the execution of a procedure (rather than at the level of a specific item) is arguably more appropriate. For the item-general practice group, the AA algorithm was practiced on every trial, and thus the functional lag was shorter than in the item-specific practice group involving memory retrieval of a specific item. Thus, on closer analysis, spacing effects do not appear to map cleanly onto the current pattern and do not provide a straightforward account of the differential gain and loss involved with item-general and item-specific practice.

Of secondary interest, potentiation was also examined in all three experiments, given the mixed results in earlier research. Although we did not find strong evidence for potentiation in any of our current experiments, response times for practiced items during the test session of Experiment 3 were better predicted by the extrapolation of Session 1 data when novel items were absent than when novel items were present, which offers a potential answer to why some data sets show evidence of potentiation and others do not. Another possible explanation may involve the types of practice gain accrued in different practice sessions. For example, in Rickard's (2007) pound arithmetic task, the algorithm route was likely involved in a greater proportion of responses during the first session than during the second session. If so, potentiation may actually reflect averaging more of the faster retrieval-based response times with slower algorithm-based response times in Session 2 than in Session 1. However, this explanation still does not account for potentiation observed in other data sets. For example, in Rickard's paired-associate task, responses would most likely be due to item-specific memory retrieval in all sessions. Therefore, potentiation would not reflect averaging different proportions of retrieval-based and algorithm-based responses in different practice sessions. Hence, further investigation is required to ascertain why potentiation occurs for some cognitive tasks and not others.

Implications for Theories of Cognitive Skill Acquisition

The current research revealed novel empirical patterns that provide important constraints for theories of cognitive skill acquisition. Complete theories of skill acquisition will need to explain not only why gains are made during practice but also why different types of practice gain are lost at different rates across delays. Currently, theories of cognitive skill acquisition and memory-based theories of automaticity have varied success with meeting these criteria.

Whereas memory-based theories of automaticity provide relatively straightforward explanations for why item-specific gains are made during practice (Logan, 1988; Palmeri, 1997; Rickard, 1997), only one memory-based model explicitly includes a parameter to allow for improvement in the item-general algorithm route (component power laws [CMPL] theory, proposed by Rickard, 1997). More important, none of the memory-based theories include assumptions concerning loss of practice gains across a delay. On the one hand, adding a forgetting parameter to these formal models would seem to be a straightforward way to accommodate loss of item-specific gains in these theories. On the other hand, although these models were not originally intended to address item-general gains, simply adding a forgetting parameter would still leave these models incomplete without explanations for differences between item-general and item-specific gains and loss.

General theories of cognitive skill acquisition, most notably ACT-R (Anderson & Lebiere, 1998), are better equipped to explain the findings from the current research. First, ACT-R can account for both item-general and item-specific practice gains. Concerning item-general practice gains, ACT-R instantiates the algorithm as a sequence of productions. For example, Johnson, Wang, and Zhang (1998) begin the AA algorithm with the production:

> If the goal is to do an alphabet arithmetic problem, but the answer has not been determined,
>
> Then set a subgoal to compute the answer by counting.

The computation subgoal then leads to the execution of a production that retrieves one of seven chunks (ABCD, EFG, HIJK, LMNOP, QRS, TUV, WXYZ) from declarative memory. Once a chunk is retrieved, the next step involves iterative execution of a production that starts at the first letter in the chunk and moves forward one step at a time until the starting letter in the AA problem is located. Then, another production is executed iteratively to move forward in the chunk the number of times designated by the addend in the AA problem. Once the terminal letter is reached, a final production compares it with the ending letter in the AA problem to determine whether the answer is true or false. As described in the introduction, repeated successful use of each of these productions will increase their strength and, in turn, the likelihood of selection and speed of execution on subsequent processing trials.

ACT-R can also account for item-specific practice gains in a manner that is functionally equivalent to the memory-based theories. For example, one way ACT-R instantiates the item-specific memory-based processing assumed by automaticity theories is via a production that retrieves a specific declarative chunk (Johnson et al., 1998):

> If the goal is to do an alphabet arithmetic problem of the form letter1 + number =, but the answer has not been determined,
>
> And there is a fact in memory stating that letter1 + number = letter2,

Then report letter2 as the answer.

If the memory-retrieval production leads to a correct response, it accrues strength. Strength determines both the likelihood of selection and the speed of the production on subsequent trials, which would contribute to item-specific gains. Additionally, the declarative chunk retrieved on any given trial will accrue strength, which will similarly improve retrieval speed on subsequent trials involving that particular item.

A second way in which ACT-R differs from memory-based theories of automaticity is that it can also account for the loss of practice gains, through the loss of strength (both for productions and for declarative chunks). The strength of an element (i.e., a production or chunk) is modeled as the sum of the individual strengthenings from each processing event involving that element, by the strength accumulation equation (Anderson, 1982; Anderson et al., 1999; Anderson & Lebiere, 1998):

$$\text{strength} = \sum_{j=1}^{n} t_j^{-d},$$

where t_j is the time since the j^{th} strengthening event, and the sum is for the total number (n) of strengthening events for a particular element. The d parameter is the decay rate for each individual strengthening event. Therefore, the overall loss of an element's strength is modeled in ACT-R by the decay of strength for each individual strengthening event.

To account for loss across a delay, Anderson et al. (1999) added parameters to the strength equation just mentioned. Specifically, they measured time as $t_j = x + mH$, where x is the number of practice blocks/repetitions an item has accrued by the end of practice, m equals the number of days between practice sessions, and H equals the number of practice blocks equivalent to a 1-day interval. Anderson et al. (1999) further conceptualized t in terms of psychological time rather than clock time, with forgetting reflecting the effect of subsequent interfering events rather than the passage of time per se. Separate m and H parameters are included to reflect greater interference from highly similar events during a practice session than from less related events in the interval between sessions.

Thus, ACT-R is currently equipped to explain loss of practice gains across a delay, but the way in which ACT-R would account for differential loss of item-general versus item-specific gains is less clear. How might this model be modified to account for differential loss? Although ACT-R models typically include the same d parameter for both productions and chunks, one possibility is that production strength decays more slowly than does declarative chunk strength. As illustrated earlier, the AA algorithm is much more heavily dependent on productions than on retrieval of declarative chunks, and thus most item-general gains come from improvements in production strength. In contrast, the item-specific retrieval route illustrated earlier involves only one production, so the proportional contribution of declarative chunk strength to response time is larger. Thus, greater decay for chunks than for productions could yield greater loss of item-specific versus item-general gains. Consistent with this possibility, Anderson et al.'s (1999) best fit d parameter was greater for their fact-retrieval task involving item-specific gains ($d = 0.76$) than for their rule application task involving more item-general gains ($d = 0.52$). Of course, chunk retrieval is also involved in the item-general algorithmic route, as outlined earlier. However, base-level activation is likely greater for the alphabet declarative chunks (e.g., ABCD) involved in the item-general algorithm than for AA declarative chunks (e.g., A + 2 = C) involved in the item-specific retrieval route, given preexperimental exposure to alphabet chunks but not to AA chunks.

Another possibility is that the passage of psychological time or number of interfering events between sessions (H) is different for productions than for chunks. If so, ACT-R could be modified to include different H parameters for productions and chunks. However, Anderson et al.'s (1999) best-fit H parameter was smaller for the fact-retrieval task involving item-specific gains ($H = 4.55$) than for the rule-application task involving more item-general gains ($H = 7.02$). These parameter estimates suggest greater susceptibility to interfering events across the delay for item-general versus item-specific gains, which runs contrary to the current pattern of greater preservation of item-general gains. Of course, the models reported by Anderson et al. were not intended to account for differences in item-general versus item-specific effects, and thus these comparisons are speculative at this point. Clearly, further model testing will be needed to identify how ACT-R may be modified to account for the difference between item-general and item-specific gains and loss.

Alternatively, another explanation for why differential loss occurs for item-general versus item-specific practice gains may not require separate parameters for productions and chunks. Rather, differential loss may be due to different amounts of functional practice for productions and chunks in the item-general and item-specific practice groups. In Experiment 2, one might assume for the sake of simplicity that participants in the item-general practice group used the algorithm for all 288 trials during practice. In contrast, one might assume that participants in the item-specific practice group retrieved each item from memory a maximum of 47 times (assuming algorithm is used for only the first presentation of each AA problem). Therefore, differential loss might appear to be due to more practice for the algorithm (288 trials) than for specific item retrieval (47 trials per item). However, consideration of functional practice at the grain size of the ACT-R elements involved for the algorithm and memory-retrieval routes reveals that differences in functional practice may be relatively minimal. For the item-general practice group, the productions of the ACT-R algorithm outlined earlier are at most practiced for 288 trials. However, for the item-specific group, the production of ACT-R memory-retrieval route outlined earlier is practiced on up to 282 trials. Similarly, for the item-general practice group, given the way that we structured our list of items, each of the seven alphabet chunks would have been retrieved on approximately 48 trials on average. For the item-specific practice group, any one of the six AA chunks could have been retrieved on up to 47 trials. Thus, functional practice differences in the item-general and item-specific practice groups are unlikely to completely account for differential loss across delays.

In sum, the present research reports the first empirical evidence concerning loss of item-general practice effects across a delay, supporting the idea that complete theories of cognitive skill acquisition need to account for the loss of practice gains across a delay. The present findings also suggest that theories of cognitive skill acquisition will require further development to be able to explain the differential effects of delays on item-general and item-specific practice effects.

References

Anderson, J. R. (1982). Acquisition of cognitive skill. *Psychological Review, 89,* 369–406. doi:10.1037/0033-295X.89.4.369

Anderson, J. R., Fincham, J. M., & Douglass, S. (1999). Practice and retention: A unifying analysis. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25,* 1120–1136. doi:10.1037/0278-7393.25.5.1120

Anderson, J. R., & Lebiere, C. (1998). *The atomic components of thought.* Mahwah, NJ: Erlbaum.

Bajic, D., & Rickard, T. C. (2009). The temporal dynamics of strategy execution in cognitive skill learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35,* 113–121. doi:10.1037/a0013647

Brainerd, C. J., Reyna, V. F., & Kneer, R. (1995). False-recognition reversal: When similarity is distinctive. *Journal of Memory and Language, 34,* 157–185. doi:10.1006/jmla.1995.1008

Cepeda, N. J., Pashler, H., Vul, E., Wixted, J. T., & Rohrer, D. (2006). Distributed practice in verbal recall tasks: A review and quantitative synthesis. *Psychological Bulletin, 132,* 354–380. doi:10.1037/0033-2909.132.3.354

Compton, B. J., & Logan, G. D. (1991). The transition from algorithm to memory retrieval in memory-based theories of automaticity. *Memory & Cognition, 19,* 151–158.

Gaschler, R., & Frensch, P. A. (2007). Is information reduction an item-specific or an item-general process? *International Journal of Psychology, 42,* 218–228. doi:10.1080/00207590701396526

Grant, S. C., & Logan, G. D. (1993). The loss of repetition priming and automaticity over time as a function of degree of initial learning. *Memory & Cognition, 21,* 611–618.

Johnson, T. R., Wang, H., & Zhang, J. (1998). Modeling speed-up and transfer of declarative and procedural knowledge. In *Proceedings of the Twentieth Annual Conference of the Cognitive Science Society* (pp. 531–536). Hillsdale, NJ: Erlbaum.

Koch, I., & Allport, A. (2006). Cue-based preparation and stimulus-based priming of tasks in task switching. *Memory & Cognition, 34,* 433–444.

Kolers, P. A. (1976). Reading a year later. *Journal of Experimental Psychology: Human Learning and Memory, 2,* 554–565. doi:10.1037/0278-7393.2.5.554

Logan, G. D. (1988). Toward an instance theory of automatization. *Psychological Review, 95,* 492–527. doi:10.1037/0033-295X.95.4.492

Logan, G. D. (1992). Shapes of reaction-time distributions and shapes of learning curves: A test of the instance theory of automaticity. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18,* 883–914. doi:10.1037/0278-7393.18.5.883

Logan, G. D., & Bundesen, C. (2003). Clever homunculus: Is there an endogenous act of control in the explicit task cuing procedure? *Journal of Experimental Psychology: Human Perception and Performance, 29,* 575–599. doi:10.1037/0096-1523.29.3.575

Logan, G. D., & Etherton, J. L. (1994). What is learned during automatization? The role of attention in constructing an instance. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20,* 1022–1050. doi:10.1037/0278-7393.20.5.1022

Logan, G. D., & Klapp, S. T. (1991). Automatizing alphabet arithmetic: I. Is extended practice necessary to produce automaticity? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 17,* 179–195. doi:10.1037/0278-7393.17.2.179

McAndrews, M. P., & Moscovitch, M. (1990). Transfer effects in implicit tests of memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16,* 772–788. doi:10.1037/0278-7393.16.5.772

Meiran, N. (1996). Reconfiguration of processing mode prior to task performance. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22,* 1423–1442. doi:10.1037/0278-7393.22.6.1423

Palmeri, T. J. (1997). Exemplar similarity and the development of automaticity. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23,* 324–354. doi:10.1037/0278-7393.23.2.324

Rawson, K. A. (2004). Exploring automaticity in text processing: Syntactic ambiguity as a test case. *Cognitive Psychology, 49,* 333–369. doi:10.1016/j.cogpsych.2004.04.001

Rawson, K. A., & Middleton, E. L. (2009). Memory-based processing as a mechanism of automaticity in text comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35,* 353–370. doi:10.1037/a0014733

Rawson, K. A., & Touron, D. R. (2009). Age differences and similarities in the shift from computation to retrieval during reading comprehension. *Psychology and Aging, 24,* 423–437. doi:10.1037/a0016044

Reyna, V. F., & Kiernan, B. (1994). Development of gist versus verbatim memory in sentence recognition: Effects of lexical familiarity, semantic content, encoding instructions, and retention interval. *Developmental Psychology, 30,* 178–191. doi:10.1037/0012-1649.30.2.178

Rickard, T. C. (1997). Bending the power law: A CMPL theory of strategy shifts and the automatization of cognitive skill. *Journal of Experimental Psychology: General, 126,* 288–311. doi:10.1037/0096-3445.126.3.288

Rickard, T. C. (2007). Forgetting and learning potentiation: Dual consequences of between-session delays in cognitive skill learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33,* 297–304. doi:10.1037/0278-7393.33.2.297

Rickard, T. C., & Bourne, L. E., Jr. (1995). An identical-elements model of basic arithmetic skills. In A. F. Healy & L. E. Bourne, Jr. (Eds.), *Learning and memory of knowledge and skills: Durability and specificity* (pp. 255–281). Newbury Park, CA: Sage.

Rickard, T. C., Lau, J. S.-H., & Pashler, H. (2008). Spacing and the transition from calculation to retrieval. *Psychonomic Bulletin & Review, 15,* 656–661. doi:10.3758/PBR.15.3.656

Rogers, R. D., & Monsell, S. (1995). Costs of a predictable switch between simple cognitive tasks. *Journal of Experimental Psychology: General, 124,* 207–231. doi:10.1037/0096-3445.124.2.207

Rubin, D. C., Hinton, S., & Wenzel, A. (1999). The precise time course of retention. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25,* 1161–1176. doi:10.1037/0278-7393.25.5.1161

Rubin, D. C., & Wenzel, A. (1996). One hundred years of forgetting: A quantitative description of retention. *Psychological Review, 103,* 734–760. doi:10.1037/0033-295X.103.4.734

Schneider, W., & Shiffrin, R. M. (1977). Controlled and automatic human information processing: I. Detection, search, and attention. *Psychological Review, 84,* 1–66. doi:10.1037/0033-295X.84.1.1

Touron, D. R., & Hertzog, C. (2004). Distinguishing age differences in knowledge, strategy use, and confidence during strategic skill acquisition. *Psychology and Aging, 19,* 452–466. doi:10.1037/0882-7974.19.3.452

Touron, D. R., Swaim, E. T., & Hertzog, C. (2007). Moderation of older adults' retrieval reluctance through task instructions and monetary incentives. *Journal of Gerontology: Psychological Sciences, 62*B, 149–155.

Wixted, J. T., & Carpenter, S. K. (2007). The Wickelgren power law and the Ebbinghaus savings function. *Psychological Science, 18,* 133–134. doi:10.1111/j.1467-9280.2007.01862.x

Yeung, N. (2010). Bottom-up influences on voluntary task switching: The elusive homunculus escapes. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 36,* 348–362. doi:10.1037/a0017894

Received August 28, 2009
Revision received March 31, 2010
Accepted April 15, 2010 ■