Practice can yield remarkable levels of achievement but little generalization. For example, Ericsson and Chase (1982) reported the illustrative case of a college student who, following many hours of practice on a digit-span task (an often used measure of short-term memory [STM]), could successfully recall over 80 randomly ordered digits. However, the individual was limited to a typical STM span of about seven items for other, even closely related, memoranda. The specificity of the improvements observed in that and other classical studies of skill acquisition and expertise have led many to conclude that the benefits of practice on a given task do not generally extend into other realms of performance (Chase & Simon, 1973; Engle & Bukstel, 1978). By this account, although individuals may exhibit innate differences in certain domain-general capacities, training strategies seeking to promote superior performance through influence on these general capacities would be destined to fail.

Several recent studies, however, have invigorated an interest in the plausibility of using repetitive mental exercise to enhance one domain-general ability—working memory (WM)—and, in so doing, to concurrently improve performance in other cognitive skills (Jaeggi, Buschkuehl, Jonides, & Perrig, 2008; Klingberg et al., 2005; Persson & Reuter-Lorenz, 2008). Training regimens used in these and other recent efforts appear to have targeted domain-general processes that individuals utilize to broadly support complex cognition. For example, in a study by Verhaeghen, Cerella, and Basak (2004), WM training was shown to influence recall from WM by expanding the capacity of attention. Likewise, others have demonstrated that WM training can impact domain-general cognitive control mechanisms (Klingberg et al., 2005; Klingberg, Forssberg, & Westerberg, 2002), interference resolution processes (Persson & Reuter-Lorenz, 2008), WM updating processes (Dahlin, Neely, Larsson, Bäckman, & Nyberg, 2008), and even general fluid intelligence (Jaeggi et al., 2008).

Although it provides a very promising foundation, the small corpus of existing WM training studies is limited in two important ways. First, prior demonstrations of transfer included measures closely related to those used to estimate WM itself, rather than to more distant tasks. To address this limitation, we sought to examine a broader battery of measures with the expectation that other tasks known to correlate with individual differences in WM capacity might also benefit from WM training. Second, previous studies utilized training tasks that are not commonly employed in the basic behavioral or psychometric literatures (e.g., atypical variants of the n-back task in Verhaeghen et al. [2004] and Jaeggi et al. [2008]; a battery of video-game-like tasks, such as that used in Klingberg et al., 2005). Thus, previous findings are somewhat disjointed from the larger behavioral literature and offer only limited insights into the specific WM mechanisms influenced by training (e.g., encoding, strategic processing, updating, etc.).

Encouraged by findings regarding the malleability of WM, in the present study, we test a novel approach to increasing WM capacity through repeated practice with an
adaptive complex WM (CWM) span task and examine the generalizability of the resulting WM improvements into the broader landscape of cognitive ability. By training participants using a CWM span task, we hoped to bridge the WM training literature to the wider body of empirical work that implicates WM in complex cognition. This variety of WM task has served as the cornerstone of the psychometric (individual differences) literature on WM for over 30 years and is arguably the most reliable and widely studied predictor of complex cognition (Conway et al., 2005). The hallmark characteristic of CWM tasks is the interweaving of storage (of the test items) and processing (the secondary decision-making task) components, which places a premium on the participant’s ability to coordinate maintenance in the face of a concurrent distraction.

We anticipated that, because CWM tasks place a strong demand on mechanisms linked to domain-general attention control (Engle & Kane, 2004), a training paradigm built around this task would result in both increases of WM span and more far-reaching benefits. Accordingly, the aims of the present study were twofold. First, we sought to examine the impact of a novel training paradigm, an adaptive version of a CWM span task. Second, we sought to expand the known boundaries of generalization by administering a battery of cognitive skills assessments.

**METHOD**

**Participants**

Forty-two Temple University undergraduates (25 female) completed the study and were compensated monetarily. Trained participants had a mean age of 20.1 years, and control participants had a mean age of 20.6 years. All participants were native English speakers with no prior events or illnesses that we expected to have an impact on WM performance.

**Cognitive Assessments**

A battery of computerized and handwritten tests assessed participants’ cognitive skill levels before and after WM training. The battery of measures was chosen for its previously documented relationship to individual differences in WM ability (see, e.g., Kane et al., 2004). Temporary memory measures included verbal and spatial STM and verbal and spatial CWM span. The verbal and spatial STM tasks assessed immediate short-term recall of serially presented (1-sec stimulus, 1-sec intertask interval) letters and locations, respectively. A schematic of each CWM measure is shown in Figure 1. The verbal CWM task also involved recall of letters, but each to-be-remembered letter was preceded by 4 sec of repeated, participant-paced, lexical decisions. The spatial CWM task used the same timing parameters but tested memory for locations presented with interleaved symmetry decisions. Participants were given unlimited time to attempt recall, and, after each trial, participants received feedback indicating both the number of correctly recalled items and the overall accuracy of decision making on the intervening distractor task. For each temporary memory measure, participants were first asked to attempt recall of three test items. Following two successful trials at a given length, the number of test items was increased by one. The number of test items increased until participants failed to correctly recall two successive trials at a given list length. A participant’s span was then defined as the maximum list length for which all items were recalled in the correct serial order.

A logical reasoning assessment included two tests of verbal reasoning (ETS’s Inference and Nonsense Syllogisms tests) and two tests of spatial reasoning (ETS’s Surface Development and Paper Folding tests). Cognitive control was measured using the Stroop Color–Word Interference Test (3 blocks of 60 trials each, 50% incongruent). General fluid intelligence was measured with Raven’s Advanced Progressive Matrices (APM), using only the odd-

---

**Figure 1.** Schematic depiction of the training tasks. Each of the training sessions included 16 trials of the verbal complex working memory (CWM) task and 16 trials of the spatial CWM task. Participants began each session with a span of four test items. One additional test item was presented following 2 consecutive successful trials, and one fewer item was presented following 2 consecutive unsuccessful trials.
A first step in analysis was to track participants’ CWM span improvements over the 20 training sessions. Mean session-to-session performance over the 20 sessions indicated a steady rate of improvement across the 4 weeks of training for both verbal and spatial CWM, as shown in Figure 2.

Analyses next focused on comparisons of participants’ cognitive skill levels before and after the training interval for the following measures: temporary memory (verbal CWM, spatial CWM, verbal STM, spatial STM), ETS reasoning tasks (inference, nonsense syllogisms, paper folding, surface development), cognitive control (Stroop), general fluid intelligence (Raven’s APM), and reading comprehension (Nelson–Denny). Through these comparisons, we sought to answer two central questions about the impact of WM training: (1) Does WM training lead to an increase in the capacity of temporary memory? and (2) Do the benefits of WM training transfer to specific untrained tasks?

Since the temporary memory and ETS reasoning assessments each included performance on multiple tests, we created composite scores for each participant. The composite was calculated by normalizing the first and second assessment scores relative to first assessment performance and then averaging the normalized scores for each of the tasks in the composite. For all assessed measures, paired-samples t tests were used to compare first and second assessment scores within trained and control groups, and an independent-samples t test was used to evaluate the difference in the improvements attained by trained versus untrained participants (Table 1). We entered the study with the directional hypothesis that training would improve test scores; thus, all t tests were one-tailed.

For the temporary-memory composite, both control and trained participants showed significant test–retest improvements (i.e., significantly increased performance on the sec-
second assessment). It is important to note that the magnitude of the improvement was significantly larger among trained participants, as confirmed by an independent-samples *t* test contrasting the size of the training effect (i.e., difference in performance from first to second assessment) for trained versus control participants (Table 1). A repeated measures ANOVA was used to further establish the strong influence of training on temporary memory capacity and revealed a highly significant interaction between assessment day and training group \( [F(1,40) = 7.483, p < .005, \eta_{p}^2 = .428] \). Since training was composed of verbal and spatial CWM tasks, we further examined trained participants’ performance on these specific measures. Groupwise gains were statistically significant for each measure \( [t(19) > 4.64, p < .005] \), and, of the 20 trained participants, 18 showed increased verbal CWM performance, and 15 showed increased spatial CWM performance.

Having found a significant impact of training on WM, we next examined transfer from training to measures that assessed more disparate cognitive abilities. Since generalization presupposes improvement on the trained tasks, we considered generalization effects both in the entire group of 20 trained participants and in the subgroup of 15 “successfully” trained participants, who demonstrated improved performance on spatial CWM, the more stringent of the two WM measures.

On our measure of fluid intelligence (Raven’s APM), no significant improvement was recorded in either the trained group (pretraining, \( M = 12.2, SE = 0.66 \); posttraining, \( M = 12.2, SE = 0.65 \)) or the control group (pretraining, \( M = 12.4, SE = 0.72 \); posttraining, \( M = 12.2, SE = 0.57 \)) (see Table 1). Examining only the subgroup of successfully trained participants, we similarly found no improvement in Raven’s APM performance (pretraining, \( M = 13.4, SE = 0.51 \); posttraining, \( M = 13.2, SE = 0.55 \)). Thus, training does not appear to have affected performance on this measure.

Results for the ETS reasoning composite are also included in Table 1. Trained participants exhibited small, but statistically significant, improvements (pretraining, \( M = -0.19, SE = 0.12 \); posttraining, \( M = 0.20, SE = 0.13 \)). However, control participants’ test–retest improvements (pretraining, \( M = 0.17, SE = 0.14 \); posttraining, \( M = 0.37, SE = 0.15 \)) were also statistically significant. Despite slightly larger gains among trained participants, an independent-samples *t* test directly contrasting improvements for the trained versus control groups was non-significant. Gains observed on these reasoning measures may, thus, reflect only test–retest improvements and not a true effect of transfer from training. This interpretation is further supported by the finding that ETS reasoning improvements were no larger for the subset of successfully trained participants (pretraining, \( M = -0.18, SE = 0.15 \); posttraining, \( M = 0.16, SE = 0.16 \)).

By contrast, for both Stroop and reading comprehension performance, trained participants exhibited improvements following training that exceeded the improvements found

### Table 1

<table>
<thead>
<tr>
<th>Measure</th>
<th>Control Group Test–Retest</th>
<th>Trained Group Test–Retest</th>
<th>Differences Between Groups</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td><em>t</em></td>
<td>df</td>
<td><em>p</em> Value</td>
</tr>
<tr>
<td>Temporary memory</td>
<td>1.96</td>
<td>21</td>
<td>.032</td>
</tr>
<tr>
<td>Cognitive control</td>
<td>1.81</td>
<td>21</td>
<td>.043</td>
</tr>
<tr>
<td>Reading comprehension</td>
<td>0.87</td>
<td>20</td>
<td>.200</td>
</tr>
<tr>
<td>ETS reasoning battery</td>
<td>2.09</td>
<td>21</td>
<td>.024</td>
</tr>
<tr>
<td>Fluid intelligence</td>
<td>0.31</td>
<td>21</td>
<td>.380</td>
</tr>
</tbody>
</table>

![Figure 3](image-url) Pre- and posttest performance in the Stroop and Nelson–Denny reading comprehension measures for matched trained and untrained participants. The Stroop congruency effect is inverted to better illustrate training gains.
in control participants (Figure 3). Both control and trained participants showed significant test–retest improvements in Stroop performance (i.e., reduction of the Stroop congruency effect). An independent-samples r test of the change scores indicated significantly greater improvement among trained participants than among controls (Table 1), although a repeated measures ANOVA did not yield a significant training group (trained, control) × assessment day (pretraining, posttraining) interaction \( F(1,40) = 2.39, p = .13, \eta^2 = .056 \). Still, confidence in the impact of training on Stroop performance is encouraged by the finding that reductions in the Stroop congruency effect were more substantial among the subgroup of successfully trained participants, and a repeated measures ANOVA based on this training subgroup produced a marginally significant interaction between training group (successfully trained, control) and assessment day (pretraining, posttraining) \( F(1,35) = 4.03, p = .053, \eta^2 = .103 \).

Effects of training on reading comprehension are summarized in Table 1 and Figure 3. Examination of pre- and posttraining reading comprehension performance showed significant test–retest improvements for trained participants, but not for control participants. A direct contrast of reading comprehension improvements in trained versus untrained groups was also significant. Repeated measures ANOVAs showed a marginally significant interaction between training group and assessment day when data from the entire training group were considered \( F(1,38) = 3.30, p = .077, \eta^2 = .08 \); and, again, transfer from training was more strongly indicated in the subset of successfully trained participants \( F(1,34) = 6.94, p = .013, \eta^2 = .174 \).

Our findings, which suggest an effect of WM training on Stroop and reading comprehension performance, can be informed further by a closer examination of individual differences in the influence of training on WM capacity itself. As we have already shown above, participants for whom training promoted an increase in WM span (i.e., successfully trained participants) exhibited stronger improvements for each transfer measure. Taking a more nuanced view, one might further expect to find a correlation between WM improvements and gains on the transfer measures, with greater improvements in those for whom the WM training was most effective (i.e., produced larger increases in WM span) and weaker improvements in those for whom it was relatively less effective. Despite our relatively small sample size, we accordingly tested correlations between WM span increases and both Stroop and reading comprehension improvements. For the Stroop test, the correlations were weak \( r(18) = -.01 \) for verbal WM; \( r(18) = .12 \) for spatial WM and not significant. However, the results from the reading comprehension measure were more encouraging. The correlation between verbal WM increases and Nelson–Denny improvements was modest \( r(18) = .24, p = .13 \), and there was a strong and statistically significant relationship between trained participants’ spatial CWM span increases and reading comprehension improvements \( r(18) = .49, p < .005 \), as is shown in Figure 4. Among control participants, there was no observed relationship between changes in WM span and Stroop performance \( r(22) = .19 \) for verbal WM; \( r(22) = -.10 \) for spatial WM] or between changes in CWM span and reading span \( r(20) = -.28, p = .90 \) for verbal CWM; \( r(20) = .097, p = .67 \) for spatial CWM.

**DISCUSSION**

Our objectives in the study were twofold. First, we assessed the efficacy of a novel WM training paradigm, an adaptive version of a CWM span task that could serve as a bridge between the WM training and psychometric literatures. Consistent with this aim and with earlier studies (Jaeggi et al., 2008; Klingberg et al., 2005; Klingberg et al., 2002; Verhaeghen et al., 2004), we found that trained participants improved significantly more than did controls on temporary memory measures.

Second, we aimed to examine the scope of transfer from training. If training influences specific processes shared only between a particular cognitive measure and the trained tasks, then narrow generalization would be expected. On the other hand, transfer to multiple (disparate) measures could be taken as evidence that training has an impact on a domain-general mechanism.

Others have shown that WM training benefits are not task specific, but rather, that they extend beyond the trained task by affecting WM processes, including interference buffering (Persson & Reuter-Lorenz, 2008) and WM updating (Dahlin et al., 2008). A few tantalizing studies have further demonstrated transfer of WM training to other assessments of cognition, including measures of cognitive control (Klingberg et al., 2005) and fluid intelligence (Jaeggi et al., 2008).

We administered a battery of cognitive skill assessments in order to determine whether WM training would afford even broader generalization than has been previously demonstrated. Generalization was not universally indicated: Some of the skills we assessed did not improve...
differentially with WM training (e.g., reasoning tasks, Raven’s APM). However, our results replicate the finding that WM training can lead to improvements in cognitive control, as measured in the Stroop task (Klingberg et al., 2005). In a novel finding, we further demonstrate that WM training can lead to improvements in reading comprehension. As we discuss below, the observation of generalization to two highly disparate cognitive measures (Stroop and Nelson–Denny) leads us to favor the view that WM training can indeed have an impact on a domain-general mechanism.

Of course, CWM span tasks place demands on many different processes, including encoding, covert maintenance, attention, updating, interference resolution, and controlled memory search (and perhaps many others). Training may influence one or more of these processes. Accordingly, transfer to different cognitive measures may depend on the impact of training on different processes. For example, training may improve Stroop performance by strengthening controlled attention, but it may improve reading comprehension by increasing the capacity of verbal storage.

The latter view may have particular intuitive appeal, since increased verbal storage would presumably allow participants to maintain an expanded verbatim representation of recent text and to thus integrate and assimilate the meaning of the text more easily. However, if increased verbal storage capacity explains Nelson–Denny reading comprehension increases, then individual improvements in verbal WM should strongly predict the magnitude of improvements in reading comprehension. Instead, we found that verbal WM improvements were only a weak and nonsignificant predictor of reading comprehension gains. Meanwhile, spatial WM increases were a much stronger predictor.

These findings can be addressed from the more parsimonious view that CWM training affects reading comprehension in the same way that it benefits both WM (verbal and spatial) and Stroop performance: by enhancing a domain-general attentional mechanism not specifically tied to verbal or spatial storage per se, but to the coordination of information maintenance in the face of additional processing demands. This interpretation is consistent with the broader claim that CWM tasks are highly predictive of cognitive function specifically because they limit rehearsal and domain-specific strategy use and, hence, emphasize the role of domain-general attentional mechanisms (Cowan et al., 2005). Since verbal rehearsal strategies are so highly developed, the suppression of rehearsal and the isolation of attention control mechanisms may be more complete in spatial CWM span tasks (Conway et al., 2005). Accordingly, the inclusion of a spatial CWM component in training may have been especially important in producing the observed generalization effect. Indeed, our correlational findings suggest that reading comprehension improvements were particularly dependent on increased spatial WM performance.

Admittedly, one difficulty for the claim that transfer arises from the impact of WM training on a domain-general attentional process is that both improved and unimproved measures included in our assessment battery have been linked, on the basis of latent variable studies, to this shared WM mechanism (Kane et al., 2004). The selectivity of transfer invites the question of why training generalized to some of the cognitive skills that we measured, but not to others. Specifically, the training group did not show differential improvement on measures of reasoning or fluid intelligence, which also putatively demand controlled attention. One way to address the apparent discrepancy is to assume that tasks for which we observed improvements simply place a higher demand on controlled attention than do the unaffected tasks. Another possibility is that our study lacked the statistical power to reveal modest training effects in seemingly unaffected measures. Indeed, there was a trend toward significance in the effect of training on the ETS reasoning composite that would likely have reached statistical significance in a larger sample.

However, such power limitations do not readily account for our failure to replicate a transfer of WM training benefits to measures of fluid intelligence (as was observed by Jaeggi et al., 2008), since we did not find even a trend for improvement in trained participants on Raven’s APM. Beyond statistical explanations, differences in the training paradigms used for the two studies may explain the differences in transfer effects. The training program used by Jaeggi et al. (2008) involved 400 trials per training session, with a dual n-back training paradigm designed to emphasize binding processes and task management. Conversely, our training paradigm included only 32 trials per session and more heavily emphasized maintenance in the face of distraction. Finally, the seemingly conflicting results may be due to differences in intelligence test administration. As was pointed out in a recent critique (Moody, 2009), Jaeggi et al. (2008) used atypical speeded procedures in administering their tests of fluid intelligence, and these alterations may have confounded the apparent effect of WM training on intelligence.

The selective pattern of transfer that we observed is informative in another regard. As was the case in almost every previous WM training study (cf. Persson & Reuter-Lorenz, 2008), our experiment lacked an “active” control group (i.e., a group engaged in alternate activities during the training interval). We cannot, therefore, rule out that observed improvements were the result of differences in the motivation level or expectations of trained, relative to untrained, participants: a placebo effect. However, an explanation of this type does not explain the selectivity of transfer. If trained participants were simply more motivated at the time of the second assessment, differences in motivation level should have been as apparent on tests of fluid intelligence and reasoning as on tests of WM, cognitive control, and reading comprehension. Additionally, the observed correlation between spatial CWM and Nelson–Denny improvements is not readily explained as a simple placebo effect.

In sum, the present study demonstrates success in using a CWM task to improve WM, cognitive control, and reading comprehension. Isolated demonstrations of generalization from WM training should be interpreted with
some degree of caution, and further work in this area—including replications, investigations of durability, tests of relative efficacy, and considerations of the real-world benefits of different WM training paradigms—is clearly needed. Still, the present findings constitute an important step toward bridging the WM training literature and a long history of empirical and psychometric work based on CWM tasks. Moreover, the results are consistent with the hypothesis that WM training can have an impact on domain-general cognitive mechanisms; thus, these results benefit multiple areas of cognition. In particular, our discovery that WM training can yield improvements in reading skill—even among college-aged participants—encourages enthusiasm for the increasingly popular belief that WM training can be used as a general tool for promoting important cognitive skills.

AUTHOR NOTE

The authors thank Jenny Andrade for her role in data collection. Correspondence concerning this article should be addressed to J. M. Chein, Department of Psychology, Temple University, 825 Weiss Hall, 1701 N. 13th Street, Philadelphia, PA 19122 (e-mail: jchein@temple.edu).

REFERENCES


NOTES

1. The Raven’s APM is not a speeded test. All participants completed the test well within 45 min.

2. Second-assessment reading comprehension data were missing for 1 trained and 1 control participant; these participants were, therefore, necessarily excluded from analyses of the reading comprehension measure.

(Manuscript received July 17, 2009; revision accepted for publication November 25, 2009.)