Efficient cognitive operations predict skill acquisition

Francis X. Smith Jr. a, Christopher A. Was b,⁎

a University of Iowa, United States
b Kent State University, 405 White Hall, Kent, OH 44242, United States

1. Introduction

In recent years a number of theorists have noted that typical attention-driven working memory models alone cannot account for a wide range of complex cognitive tasks, such as language comprehension, in which large amounts of information must be maintained to perform the task at hand (Anderson, 1983; Cowan, 1999; Ericsson & Delaney, 1999; Just & Carpenter, 1992). Even Baddeley’s model, the original working memory model to postulate attention-driven components, now includes an episodic buffer, which recruits long-term memory elements to support cognitive activities (Baddeley, 2000). Indeed, a variety of alternative models of working memory have been proposed which include long-term memory when considering complex cognition (see Miyake & Shah, 1999). It is now widely accepted that working memory and long-term memory work together to facilitate a wide range of cognitive tasks.

One such way that working memory and long-term memory may collaborate is through the facilitation of procedural memory (PMM). Procedural memory traces are the basis for skilled performance of cognitive operations. As defined by Woltz and Was (2006, 2007), PMM may be considered as the strengthening of persistent memory for prior cognitive operations. Important to this definition is that semantic and declarative information are not the focus of PMM. It is the procedures that take place, such as searching memory or applying appropriate procedures to stimuli, which are strengthened.

Of the numerous theories that include elements of long-term memory in the performance of cognitive tasks, the most relevant is the ACT-R architecture as it provides the most direct parallel to PMM in the present study. While not strictly speaking a model of working memory, Anderson’s ACT-R model is able to explain human performance during cognitive activities that could not otherwise be explained given the conventional limits of short-term memory (Anderson, 1983). Within the ACT-R framework there are two definitions of working memory. The first definition is that working memory is primarily the content that is maintained during processing whereas the second definition is any of the processes that allow memory elements to be maintained concurrently (Lovett, Reder, & Lebiere, 1999). The contents of working memory are those declarative and procedural nodes that are more highly activated and can thus be easily accessed relative to other less activated declarative nodes. Both declarative memory and procedural memory rely on the spreading of source activation based on the goal of the cognitive system.

The ACT-R system contains multiple buffers used to coordinate the cognitive system including both a goal module and a declarative memory module, but the driving force for cognition within the framework is the production rule system (Anderson et al., 2004). The production rule system is a complex system that recognizes patterns of information presented within the buffers from the goal module, declarative memory module, and a perceptual-motor module and uses this information to select one production rule to apply based on prior experience and expected utility. As tasks are repeatedly performed, the expected utility for any production rule is altered based on prior success and failure. If a production routinely results in success it would receive greater priority in the future when similar information is loaded into the buffers from goal modules and declarative memory modules.

Although this brief presentation of the ACT-R model does not explain the calculations behind the model in depth, it relates the concepts of the model to the goals of the present study. Given the goals of predicting skill acquisition, the Stroop color-word task was
chosen to measure improvement in performance across trials. Considering repeated performance on the Stroop task, the ACT-R framework has already been used to model performance on the Stroop task (Altmann & Davidson, 2001). In Altmann and Davidson’s modeling of Stroop task performance, it is assumed that the automaticity of reading causes initial discrepancy in responses to incongruent trials. Further exposure to the Stroop task allows the ACT-R framework to adjust the utility value for the production rule for responding to the printed word in order to favor color naming over repeated trials. This adjustment in utility value for a procedure results in facilitation of the procedural memory. Put differently, as procedures result in success they are likely to be favored in future trials. Adjusting utility values over time for successful procedures strengthens prior successful cognitive operations. If the ALTM task is truly capturing a measure of content-specific but item-general FPM, it may be the case that this measure may also relate to the degree of transfer between the color-word Stroop task and the number Stroop task.

The Stroop color-naming task was first developed to measure interference in serial verbal responses (Stroop, 1935). In the years since the first report of the Stroop effect, it has received a great deal of attention in the literature and has still not been adequately explained though it is often interpreted as a subject’s ability to inhibit a prepotent response, reading the displayed word, in favor of a slower controlled response (MacLeod, 1992). It is the very automaticity of reading words that is thought to produce the conflict when subjects are asked to name the color of the ink in incongruent trials (MacLeod & MacDonald, 2000). Because the Stroop task represents a subject’s ability to maintain a goal and select an appropriate procedure, color naming, in the face of a more automatic response makes it an ideal task to evaluate facilitation of procedural memory. Increased performance, in the present study, is operationalized as a decrease in latency on incongruent trials on the Stroop task.

1.1. The ALTM task

In an attempt to measure the construct of available long-term memory, Woltz and Was (2006) developed the ALTM task. The experimental task as initially employed included four discrete components (see Fig. 1). The first was a memory load in which a subject is presented words to compare and maintain in working memory. The second component in the initial experiments was a selection instruction asking subjects to focus on or ignore one of the types of items presented. Recall of the appropriate items was a third component of the experiment. Finally, subjects were asked to make comparisons between both old and new exemplars as either similar or dissimilar. The assumptions behind the logic of this task are based on the idea that attention-driven processing in working memory should lead to greater accessibility of long-term memory elements, in concordance with a
number of working memory models (e.g., Anderson, 1983; Cowan, 1999; Ericsson & Kintsch, 1995).

If the ALTM task is measuring FPM then it should relate to a decrease in latency in incongruent trials in the Stroop task over practice. Furthermore, it should relate to this decrease in latency to a higher degree than it relates to an overall decrease in latency on congruent trials because facilitation of procedural memory should not decrease word-reading speed specifically but rather make color-naming faster.

1.1.1. Components of the ALTM task

The first component, the memory load, is intended to engage subjects’ attention-driven working memory processes (Woltz & Was, 2006). Subjects are presented with words from two different categories in succession and asked to maintain these words for later recall or comparison. In previous findings by both Cowan (1999) and Oberauer (2002) active information in long-term memory has been shown to correspond to items processed, even if not consistently maintained or manipulated, in working memory. This suggests that simple presentation of words can activate long-term memory resources even if extensive processing is not required. The claim made in the present study is not that activation of declarative memory, but rather the actual procedure of comparing items, involved in later phases, contributes to long-term memory’s involvement in this task.

The memory load phase of the task can affect the magnitude of response facilitation based on the amount and type of processing required (Woltz & Was, 2006). The initial memory load size has an effect on the amount of facilitation in later category comparison trials. When memory load is required to be maintained concurrently with category comparison operations, there was no decrement in response facilitation of related exemplars that had not been previously presented, suggesting that the processes underlying ALTM either require no further attentional resources or draw upon some resource distinct from that used to actively maintain information in working memory, such as facilitating the procedures needed to make comparisons (Woltz & Was, 2006). The type of information presented within the memory load is also important. If a subject is presented with category exemplars during the memory load, later category feature comparisons are relatively unprimed (Woltz & Was, 2007). These results suggest that it is not simple spreading of activation or general semantic priming of related information leading to facilitated response times and accuracy. Rather it is likely that the cognitive procedures carried out during the memory load and subsequent phases produce the facilitation effects observed (Was, 2010b).

Following the memory load there is typically some type of selection instruction. In this phase of the task a subject is asked to either remember one category or to ignore one category. The results of prior research have shown that the selection instruction itself has a large effect on degree of facilitation (Woltz & Was, 2006). Subjects asked to remember a category show greater facilitation for later exemplars related to that category. Interestingly, subjects asked to ignore a category also show greater facilitation for later exemplars related to the ignored category. One explanation for these results is that the cognitive processes involved in identifying the exemplars are causing facilitation effects rather than simply the maintenance of the exemplars. To successfully ignore a category of exemplars it is necessary that a subject process each exemplar presented as either a member of the to-be-ignored category or not. Even if the subject then maintains the alternative category’s words in working memory for recall, facilitation effects were shown to be greater for the ignored category (Woltz & Was, 2006).

The final component of the ALTM task is the most important. In the final phase of the task, subjects are shown exemplars from the two categories previously presented in the memory load as well as exemplars from a third category which was not previously presented. Responses to these trials are in the form of like, for words that come from the same category, or dislike, for words that are from different categories. Performance on category comparisons from groups previously presented is compared to performance with the category which has not been previously encountered to examine facilitation in both response speed and accuracy. In prior research, a combination and transformation of both response time and accuracy has been used to measure facilitation (Woltz & Was, 2006). This transformation includes useful information regarding both response speed and correct responses and is interpreted as correct responses per minute. This measure has also been shown to approximate a normal distribution more closely than either measure does individually (Woltz & Was, 2006). It is also important to note that only positive match trials were evaluated in previous research due to negligible priming effects in negative match conditions (Woltz & Was, 2006, 2007).

Support of a facilitation of procedural memory account of ALTM task outcomes is demonstrated by the duration of the facilitation effects. Was (2010b) found that facilitation effects could still be found even when the comparison phase was presented 24 h after the initial memory load phase of the task. Typical accounts of long-term semantic priming (LTSP) cannot account for these durations, but procedural memory adjustments can be potentially long-lasting (Was, 2010b). The long-lasting priming observed in past ALTM experiments could easily be due to FPM but cannot be easily explained by semantic priming. Because of this it seems most likely that FPM drives a significant portion of the effects observed in past uses of the ALTM task.

Furthermore, the significant facilitation effects demonstrated by Was (2010b) for new exemplars not encountered during the memory load in Day 1 cannot be explained by expectancy theories of priming. The participants in the Was study would have no reason to expect to see the stimuli in Day 1 or related stimuli on Day 2 as they were not informed of the nature of the study. Based on the literature reviewed, we agree that the ALTM task represents a facilitation of operation-specific but item-general memory operations.

1.1.2. Uses of the ALTM task

It is important to understand what the ALTM task is measuring because recent research utilizing the task has identified the predictive values of the task measure. In the present study, the Stroop task was used to examine the relationship between performance on the ALTM task and a relatively simple cognitive task. While the Stroop task requires substantial attention to successfully complete, it does not require a great deal of information maintenance outside of task goals. If the ALTM task only measures a form of long-term semantic priming it would be unlikely that performance on the two tasks would be related. Activation of related long-term memory elements (colors, in this case) would not be likely to increase performance at naming the colors as opposed to reading words. If, however, the ALTM task does indeed represent the subject’s facilitation of procedural memory then performance on the ALTM task and gains in performance on the Stroop task performance should be positively related. As explained by the ACT-R model, in the context of the present study performance on the ALTM task may be related to the degree to which a subject can adjust the utility value of color naming compared to reading text for the purposes of the Stroop task (Altmann & Davidson, 2001).

The importance of identifying what the ALTM task is measuring is based on the usefulness of the ALTM task in explaining variance in a number of complex cognitive activities. The ALTM task has been shown to explain significant variance in complex cognitive activities such as reading comprehension (Was, 2010a). These findings are similar to results obtained earlier by Was and Woltz (2007), which indicated that ALTM mediates the relationship of not only working memory but also background knowledge with listening comprehension. There is clearly a role for the ALTM task when considering complex cognitive activities for which working memory capacity alone simply cannot adequately account.

Perhaps the most powerful finding is that the ALTM task predicts unique variance in fluid intelligence and comprehension scores even
after accounting for working memory (Was, Dunlosky, Bailey, & Rawson, 2012). In this study, Was and colleagues used structural equation modeling to investigate the effects of both working memory and facilitation of procedural memory on measures of fluid intelligence and comprehension. Whereas facilitation of procedural memory and working memory did share variance, the ALTM task measures also predicted unique variance as well. These findings indicate that the ALTM task's measure of facilitation of procedural memory and working memory tasks are measuring two distinct constructs. The ALTM task may be measuring facilitation of procedural memory in that it represents an individual's ability to efficiently apply a procedure from memory to stimulus items in the environment (Was et al., 2012). This measure, then, should be effective in predicting a decrease in the Stroop effect across multiple sessions. It should indicate how effective an individual is at increasing color naming, rather than word reading, over time.

1.2. The Stroop task

The initial study reporting the Stroop effect was conducted over 75 years ago (Stroop, 1935). Stroop found two important effects that are relevant to this day in regard. First, in the original study, there were gender-based differences in performance on the Stroop task. Stroop interpreted these differences as representative of an increase in ability for color naming in females (Stroop, 1935).

The second important effect demonstrated by Stroop was nearly one-fourth of the increase in performance across trials is accounted for by an increase in speed of color naming (Stroop, 1935). This second finding motivates the use of the Stroop task in the present study. Facilitation of procedural memory may be responsible for individual differences in the increase in speed of color naming relative to word reading. This would not be due to more efficient inhibition as such. Because some of the increased in performance can be attributed to increased speed in color naming it is likely that this procedure is being enhanced rather than word reading simply being inhibited. If a decrease in the Stroop effect over practice was due to simple inhibition word-reading trials would be responded to more slowly with practice. If, instead, color-naming procedures are facilitated through practice the decrease in the Stroop effect would be caused by speeded responses to color-naming incongruent trials. This leads to the argument that something beyond simple inhibition is occurring. Individuals are likely adjusting the relative “weight” given to various procedures while performing the task.

The Stroop task has a long history and is only reviewed briefly here, though MacLeod (1991) provided an excellent review of the overall findings from over half a century of Stroop task research. In a paper written shortly after his review, MacLeod (1992) referred to the Stroop task as the “gold standard” of attentional measures and went on to explain that it holds such a strong position in research because it has not been adequately explained. The present study is not an attempt to explain the existence of the Stroop interference effect but rather an attempt to account for individual differences in improvement in performance on the Stroop task across repeated trials.

Although the Stroop task is not as complex in its structure as the previously discussed ALTM task, it is still worth reviewing the components of the task and their impact on the present study. The general structure of any Stroop task involves presenting a subject with stimuli that vary across two dimensions, one of which is to be ignored (Cohen, Dunbar, & McClelland, 1990). In the classic color–word Stroop task subjects are presented with a stimulus that is the written word for a color such as “red” which is either printed in the same color as the written word or a different color. When the word presented is the same as the color of the ink in which it is presented, the trial is referred to as congruent. When the word presented is written in a color that differs from the word itself, the trial is referred to as incongruent. Across all trials, subjects are instructed to respond to the color of the ink and to ignore the written word. This proves difficult for many individuals.

Incongruent trials within the color–word Stroop task produce more errors as well as longer response times than do congruent trials (MacLeod, 1991). It is this very interference that Stroop (1935) sought to investigate. The Stroop effect, the interference produced by incongruent trials, is also larger in experiments which mix incongruent and congruent trials when compared to experiments which present only incongruent trials (MacLeod, 1991). One possible explanation for the different effects of interference is an individual’s ability to maintain goal-directed information during congruent trials which allow the subject to select either the word or the color and still respond accurately (Kane & Engle, 2003). In their paper, Kane and Engle suggested that subjects with higher working memory capacity are more able to maintain the proper goal, ignoring printed words to attend to colors, in the face of larger spans of congruent trials, whereas subjects with lower working memory capacity often neglect their goal so that when an incongruent trial is presented they respond incorrectly to the printed word instead of the color.

Although goal maintenance should certainly play some role in overall Stroop performance, we feel it is not a tenable explanation for improvement in Stroop performance with practice. If, as described by Kane and Engle (2003), individual differences in goal maintenance are the source of performance differences on the Stroop task, and these individual differences are due to differences in working memory capacity, then it unlikely that brief practice would improve Stroop performance. Put differently, goal-maintenance ability is also not likely to increase across trials based on working memory capacity alone.

One suggestion specifically regarding the Stroop task interference is that the controlled process of color naming and the automatic process of word reading are actually qualitatively identical and that the primary difference between the two is the strength of the associated processing pathway (Cohen et al., 1990). This view would predict that individual differences in the ability to modify these pathways should predict improvement in Stroop performance with practice. This could also potentially account for the larger Stroop effect present in experiments that mix incongruent and congruent trials (MacLeod, 1991). The account of goal neglect (Kane & Engle, 2003) predicts that these larger effects arise because individuals may fail to maintain the appropriate goal when faced with some trials in which either color naming or word reading may result in success (congruent trials) and some trials in which only color naming will result in success (incongruent trials).

The processing pathways account (Cohen et al., 1990) on the other hand would suggest that in experiments with a mix of trials it is simply the case that on congruent trials either or both pathways may be strengthened while on incongruent trials only one pathway is strengthened. Thus in an experiment with only incongruent trials the appropriate pathway should be strengthened quickly while in a mixed trial experiment the appropriate pathway may only be intermittently strengthened.

Repeated trials of the Stroop task do tend to decrease distractor potency though the reduction in the strength of the distractor is often highly specific and based on semantic codes for the distractors (Reisberg, Baron, & Kemler, 1980). This reduction in the effects of distractors may lead to some trouble detecting the effects of individual differences in the facilitation of procedural memory. Some studies have shown, however, that practice with the Stroop task can generalize across different versions of the Stroop task (Davidson, Zacks, & Williams, 2003). These findings indicate that there may be some process-specific but content-general change in Stroop performance. The Stroop number task may be facilitated by practice on the color–word Stroop task and is included in the present study to examine this possibility.

There is a clear indication that practice is important for the Stroop task (MacLeod, 1991), and the explanations offered often involve pitting a strong processing pathway against a weak one. Whereas the Stroop task has primarily been used as a measure of controlled attention or inhibition, in the present context the primary interest is not the
interference itself but rather in an individual’s ability to reduce interference through practice. The ALTM task should not, for instance, predict initial performance on the Stroop task. It may, however, be predictive of the improvement in performance across trials due to a change in procedural memory for color naming. In the present study, the difference between congruent and incongruent trials is only of interest in that it provides a metric for measuring interference and thus allows for comparisons across trial blocks to measure improvement.

2. Method

2.1. Participants

A total of 82 undergraduate students (54 females, 23 males, 5 missing demographic information) from a large Midwestern university participated in the study. Mean age of the participants in this sample was 21 years (range, 18–47). Any participant missing data for any of the measures used in the present study was not included in analysis. This included 12 cases missing data from at least one task.

2.2. Materials

2.2.1. Apparatus

All tasks in this experiment were completed on personal computers with 17-inch SVGA monitors and standard keyboards. The Z, X, N, and M keys on each keyboard were modified for the color-word Stroop task. Each of these four keys was covered entirely with a sticker of a different color. The Z key was covered with a red sticker, the X key with a yellow sticker, the N key with a green sticker, and the M key with a blue sticker. All tasks were programmed using E-Prime software.

2.2.2. ALTM task

Synonym stimuli were adapted from Was and Woltz (2007) and were organized into groups of primed and unprimed synonyms. Similar to Was and Woltz, primed synonyms were not divided into focused and ignored categories because subjects were not asked to remember a specific synonym type and were instead asked to maintain all words presented in the initial list for a frequency judgment after the memory load component. All items and instructions were presented visually to the subjects. Prior research has demonstrated that the facilitation effects in the ALTM are not reliant on perceptual priming (Was, 2010b).

Subjects were presented with instructions for each component of the task prior to any practice trials. The instructions indicated that a list of five words was to be presented and that a question would be asked about these five words later. This was to encourage participants to attend to each of the five words individually. A sample list of five words was shown (e.g., rich, wealthy, acknowledge, disclose, and affluent) as well as a sample question regarding the frequency of types of words in the list (Are there more words that mean ‘well-to-do’ or more words that mean ‘confess’). The final instruction informed participants that they would then be asked to compare words based on the similarity of their meanings. If two words were similar, subjects were instructed to press the “L” key. If two words were dissimilar, the subject was instructed to press the “D” key. Following this, the whole task structure (word list, frequency question, word comparisons) was presented on one screen as a reminder.

After completing the instruction screens, subjects were given four practice trials on the synonym comparison task. During synonym comparisons, both in the practice and actual trials, an attention screen appeared prior to word presentation with two asterisks appearing at the two locations in which words were to appear. These asterisks were displayed on the screen for 750 ms and then replaced by the two words for comparison which remained on the screen until the subject responded with either “L” or “D.” Following the subject’s response there was a delay of 1000 ms before the asterisks and subsequent word pair was presented. Throughout the synonym comparison frames there was a reminder at the bottom of the screen that “L” was the response for similar words and “D” was the response for dissimilar words.

After the practice trials, subjects were provided with feedback about their response accuracy and their average response speed per comparison. They were then reminded that the goal was to respond as quickly as possible to synonym comparisons without making careless errors. This screen was self-paced and subjects were instructed to press the space bar key when they were ready to begin the task.

Each trial block began with the words “Get ready for a word list” presented on the screen for 2500 ms. Next, the five words in the memory load (e.g., skinny, punch, slim, trim, and pound) were presented for 1500 ms each, with a 500 ms blank screen and an attention asterisk for 500 ms in the location the word was to appear. After the fifth word was presented there was a 2500 ms delay during which the screen was blank. Following this delay the subject was asked to judge the relative frequency of two concepts from the word list. Each word list contained two synonyms from one concept group and three synonyms from the other. The three similar words from the more numerous concept group were never presented consecutively. Subjects were asked “Were there more words that meant Concept 1 (SLENDER) or more words that meant Concept 2 (TO HIT)?” and below this question both concepts were displayed with instructions to press the “1” key for concept one or the “2” key for concept two. This screen was self-paced and would not advance until the subject selected one of the two options. The correct response was distributed between being presented on the left or right side of the screen. Feedback was provided immediately following the subject’s response.

Following the frequency judgment, the words “Get ready to compare word meanings” were displayed. The procedure for synonym comparisons was identical to the practice trials as described above except for the number of comparison frames. During each trial, subjects were presented with four warm up comparisons which did not utilize any concepts from the synonyms presented in the word list. These warm up comparisons are not included in data analysis and served only to familiarize subjects with the task. Following the four warm up comparisons, subjects completed eight synonym comparisons. Comparisons included synonyms from the more numerous concept from the memory load as an example of direct facilitation (an actual item from the memory load) in both positive (skinny, trim) and negative matches (slim, evil), as well as an indirect (related item not presented in the memory load) positive match (thin, scrawny). The set also included the less numerous concept in the form of direct facilitation positive (punch, thump) and negative matches (pound, march), as well as an indirect positive match (hammer, pummel). The final two items were from a neutral concept (not presented in the memory load at all) positive (hang, dangle) and negative match (kiss, doctor). The order of trials was randomized within each block. At the end of each block, subjects received feedback on their accuracy for the entire block and their average response speed per item. Each subject completed 18 trials for a total of 144 synonym comparisons for analysis.

For this study the concept pairings in each trial were fixed for all participants in order to better examine individual differences and to eliminate the possibility of different concept pairings being responsible for observed differences between participants. The order of trials was randomized for each participant. Facilitation is operationalized as the difference in response speed and accuracy between exemplars related to the memory load and of neutral items not previously presented during the memory load.

2.2.3. Color-word Stroop task

The computerized version of the color-word Stroop task was adapted from the basic structure of Stroop’s second experiment, in which subjects must respond to the color ink a word is written in and not to the word itself, which is always a color word (Stroop, 1935). Subjects were instructed that during this task they should attempt to respond as quickly as possible for all trials while trying to avoid careless errors. On Day 1,
subjects were given a block of 16 warm up trials in which they were instructed to respond to words printed on the screen. All words were presented in white on a black background. In these trials, subjects were instructed to simply respond to the word with the appropriate colored key on the keyboard. Red was the “Z” key, yellow was the “X” key, green was the “N” key, and blue was the “M” key. All keys were indicated by colored stickers covering the face of each key. Subjects were told that this task was simply a test of their ability to respond to items quickly. They were told to try to become familiar with the position of response keys.

Following the 16 warm up trials, subjects were then informed that the task was going to change and that they were now to respond to the color of the ink the word was printed in regardless of what the word was. The available ink colors (and word presentations) were “RED,” “YELLOW,” “GREEN,” and “BLUE.” First, they were given a block of 32 practice trials to practice their responses before proceeding to data collection trials. This block of practice trials was included to ensure participants understood the task goal. Each trial began with a 500 ms wait followed by an attention stimulus consisting of an asterisk displayed in the center of the computer screen, where the stimulus the word was to be displayed. This asterisk was shown for 500 ms and was immediately followed by the word printed in the appropriate ink. Accuracy and latency feedback was provided after each trial. Incorrect correct trial feedback was corrective (e.g., “incorrect. The color was BLUE, you should have pressed the BLUE key.”) The practice block of trials consisted of 32 trials, 24 of which were congruent trials and 8 of which were incongruent.

After completion of the practice block, participants were informed that the experimental trials were to begin. Experimental trials followed the same format as the practice trials with the exception of the feedback. In the experimental trials, feedback consisted of the word “correct” or “incorrect” displayed on the screen for 500 ms. Each subject completed two blocks of three experimental sets of trials for a total of 192 trials, 75% of which were congruent. Therefore, on each day participants encountered a total of 168 congruent and 24 incongruent trials. There was a break between blocks that was self-paced.

2.2.4. Stroop number task

The number Stroop task consisted of two parts. Part one consisted of two blocks of 20 trials in which participants pressed a number key corresponding to a single digit presented in the center of the display. The purpose of these trials was to give participants practice using the four response keys with a single hand. Each block began with a warning which were incongruent.

Participants in the study completed the above tasks in two one-hour sessions in the educational psychology laboratory. The second session occurred approximately 24 h after the first. The interval between the two sessions was based on class scheduling and also based on the apparent durability in facilitation of procedural memory as described by Was (2010b). During the first sessions participants completed the ALTM task as well as the Stroop color–word task. During the second session participants completed a second iteration of the color–word task as well as the Stroop number task. The repeated Stroop color–word task was necessary to measure increased performance on the task over time. The Stroop number task was used to measure the generalizability of facilitation of procedural memory.

Correlation analysis was used to examine the relationship between the variables included. Table 1 presents the means and standard deviations for both response time and accuracy for each of the observed variables. For the ALTM task, only positive match trials were used in the analysis based on prior evidence from Woltz and Was (2006) that negative match trials demonstrate no noticeable facilitation effects based on condition.

According to the Kolmogorov–Smirnov test, the only observed variables that were normally distributed were all of the color-Stroop response times as well as the incongruent number Stroop response time (p < .05). All other measures seemed to violate the assumption of normality (p < .05). The Shapiro–Wilks test yielded identical results, with all color-Stroop response times as well as the incongruent number Stroop response time appearing to be normally distributed (p > .05) whereas the remaining observed variables did not appear to be normally distributed (p < .05).

These results were expected and were not problematic because the response times from the color-Stroop were the measure chosen to operationalize color-Stroop performance as is often the case for reporting Stroop results (Davidson et al., 2003). To deal with the non-normal distributions for the ALTM task a transformation was used. Table 2 includes the unique transformations used to analyze the results from the ALTM task which correspond to correct responses per minute. This measure as used previously by Woltz and Was (2006) better approximates a normal distribution in most cases and also includes relevant speed and accuracy information. The formula for calculating this transformation is as follows:

\[
\text{proportion correct} = \frac{\text{latency}_{\text{ms}}}{60,000}.
\]

Both the Kolmogorov–Smirnov and the Shapiro–Wilks tests confirmed that all three transformations presented in Table 2 appear to be normal (p > .05 for all tests). Following these analyses a theoretical model was
developed based on prior research with the ALTM task and evidence that it may, in fact, represent a measure of facilitation of procedural memory (Was, 2010b; Woltz & Was, 2006).

3. Results

3.1. Inferential statistics

Prior to any analysis of relationships between the proposed variables, it was important to verify that expected facilitation occurred for both the ALTM task and the Stroop tasks implemented in the current study. Performance on associated items on the ALTM task, items related to the memory load items that did not themselves appear in the memory load at any point, were hypothesized to best capture a measure of facilitation of procedural memory. These items required subjects to quickly and accurately identify categories for new words not previously experienced in the memory load and thus should provide a measure of how quickly a subject could adjust the utility for performing that type of categorization.

To confirm that there were differences between the various item types on the ALTM task, a one-way within subjects ANOVA was performed analyzing the speed accuracy transformations for each condition (see Table 2). Kolmogorov–Smirnov tests on the residuals from the ANOVA did not show normality (all p > .05) and thus the assumption of normality did not appear to be violated. Mauchly’s test for sphericity was not significant, though it did approach significant (p = .053). The effect of item type (neutral, memory load, or associated items) was significant, F(2, 156) = 120.82, p < .001, η²p = .61. Dunn–Sidak corrected pairwise comparisons revealed that all three conditions differed from one another (all p < .001).

Subjects performed better on comparisons involving items related to previously presented words than on comparisons involving neutral stimuli, which were unrelated to the prior memory load. Recall that the stimuli were not randomly assigned to condition because we are interested in individual differences in ALTM performance in relationship to Stroop performance. Therefore, we cannot say the results conclusively represent facilitation of procedural memory. However, with the evidence of previous experiments we are confident this is the case.

A two-way within subjects ANOVA was performed analyzing response times on the color–word Stroop task as a function of condition (congruent or incongruent) and trial block (Day 1 Block 1 or Day 2 Block 2). Because Stroop response time analyses often only use correct trials, all analyses on Stroop results were conducted only on trials in which the correct response was provided. Kolmogorov–Smirnov tests on the residuals from the ANOVA were not significant (all p > .771) and thus normality was assumed. While the main effects of trial block, F (1, 70) = 82.50, p < .001, η²p = .54, and condition, F (1, 70) = 266.93, p < .001, η²p = .79, were both significant the interaction between the two factors was as well, F (1, 70) = 33.60, p < .001, η²p = .32. This suggests that improvement across trials differed between congruent and incongruent trials.

To provide a more detailed analysis, response times were analyzed as a function of trial block for each condition type. Again, comparing Day 1 Block 1 (initial block) to Day 2 Block 2 (final block) the effect of trial block for incongruent conditions was significant, F (1, 70) = 68.38, p < .001, η²p = .49 with a mean difference of 86.32 ms (SE = 10.44) between initial block and final block performance. The effect of trial block for congruent trials was also significant, F (1, 70) = 80.48, p < .001, η²p = .54 with a mean difference of 39.67 ms (SE = 4.42) between initial and final block performance. While both conditions showed improvement across trial blocks it is clear from the significant interaction that incongruent trials showed greater improvement in response time than congruent trials. There is a reduction in Stroop interference occurring with practice. If performance on the ALTM task measures individual differences in FPM it would be expected that ALTM task performance would relate to the change in response times on incongruent trials.

A final one-way within subjects ANOVA was conducted to ensure that typically observed interference occurred with the number Stroop task as well. Response times were analyzed as a function of condition (congruent or incongruent). Again, the assumption of normality was not assumed to be violated as Kolmogorov–Smirnov tests on the residuals from the ANOVA were not significant (all p > .131). As expected, the difference between congruent and incongruent trial performance was significant, F (1, 75) = 274.16, p < .001, η²p = .79, confirming that there was Stroop-like interference in the task.

Following the confirmation that typically expected effects had indeed been observed in the present study, correlational analyses were performed to explore the relationships between ALTM task performance and improvement in color–word Stroop performance across trials. The inter-correlations among the variables are presented in Table 3.

Performance on neutral items on the ALTM task did not correlate significantly with any measure of response time on the Stroop task, as was hypothesized because neutral item performance should not capture a measure of facilitation of procedural memory. Of greater interest, associated items from the ALTM task appeared to correlate more strongly than memory items on incongruent trials but not on congruent trials. This also would be expected if associated items are effectively capturing a measure of facilitation of procedural memory.

Table 2

Descriptive statistics for ALTM task transformations (in correct responses per minute) (N = 70).

<table>
<thead>
<tr>
<th>Variables</th>
<th>M</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Memory load items</td>
<td>43.4</td>
<td>11.2</td>
</tr>
<tr>
<td>Associated items</td>
<td>47.0</td>
<td>13.2</td>
</tr>
<tr>
<td>Neutral items</td>
<td>34.6</td>
<td>8.9</td>
</tr>
</tbody>
</table>

Table 3

Correlations of observed variables (N = 70).

<table>
<thead>
<tr>
<th>Variable</th>
<th>1.</th>
<th>2.</th>
<th>3.</th>
<th>4.</th>
<th>5.</th>
<th>6.</th>
<th>7.</th>
<th>8.</th>
<th>9.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. ALTM memory</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.85**</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. ALTM associate</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.85**</td>
<td>.79**</td>
</tr>
<tr>
<td>3. ALTM neutral</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. CS Day 1 Block 1 congruent RT</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. CS Day 1 Block 1 incongruent RT</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.24*</td>
<td>.25*</td>
</tr>
<tr>
<td>6. CS Day 2 Block 2 congruent RT</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.27*</td>
<td>.38*</td>
</tr>
<tr>
<td>7. CS Day 2 Block 2 incongruent RT</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.34**</td>
<td>.41**</td>
</tr>
<tr>
<td>8. NS incongruent RT</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11. Sex</td>
<td></td>
<td></td>
<td></td>
<td>.02</td>
<td>.09</td>
<td>.08</td>
<td>.18</td>
<td>.02</td>
<td>.13</td>
</tr>
</tbody>
</table>

Note. CS = Color Stroop; RT = response time; NS = Number Stroop.

* p < .05.

** p < .01.
Therefore, to examine the relationship between improved performance on incongruent trials of the color–word Stroop task and the ALTM task, we estimated a power function to obtain the intercept and slope of each participant’s performance on the color–word Stroop task across days. To do so, we combined every two incongruent trials encountered by a participant into one trial. Recall that only reaction time for correct responses was used to measure Stroop performance. If the two contiguous incongruent trials were responded to correctly we created a mean reaction time of those two trials. If one of contiguous trials was responded to incorrectly, the other trial was used reaction time for that pair. Finally, if a participant incorrectly responded to two contiguous trials that cell was considered missing data. Combining two contiguous trials of incongruent stimuli was completed to eliminate empty cells in the curve estimation. This reduced the amount of missing data from 16% to 3%.

Means (with standard deviations in parentheses) for slopes and intercepts were $M = -11.07$ and $M = 1090.77$ (264.05) respectively. We also used the same method to estimate a curve function for the entire participant sample (see Fig. 2). The participant sample power function fit the data well, $R^2 = .69, F(1, 22) = 39.85, p < .001$. Based on the previously presented correlational analysis, we used regression analysis to determine if performance on the ALTM task would predict the improved performance on incongruent trials of the color–word Stroop task.

3.2. Regression analyses

Additional correlational analyses indicated a relationship between individual participants’ estimated slope and intercept on the color–word incongruent trials, $r = -.75, p < .001$. This was anticipated as it is often the case when participants are engaging in a new task that those whose first few trials are slow have more room for improvement and therefore a steeper slope. The correlation between the ALTM residual task measure and both the slope and intercept for color–word incongruent trials were not significant, $r = -.12$, and $r = -.20$ respectively. Again, for the same reason stated above this was anticipated. To examine the relationship between the ALTM task and improvement on the color–word task it was necessary to control for the intercept. A partial correlation was conducted in which the ALTM residual measure and the slope of the color–word task were correlated controlling for the intercept. This partial correlation was significant, $r = -.41, p < .001$. This analysis indicated that when accounting for one’s initial performance (intercept) on the color–word task the ALTM task is related to individual differences in improvement on the color–word task.

To further understand this relationship, we conducted a series of hierarchical linear regressions. In the first hierarchical linear regression, the slope of the power function across both days of the color–word Stroop task served as the dependent variable. In the first block of the regression we entered the intercept of the power function as an independent variable, then in the second block performance on the ALTM task. The full model was significant, $F(1, 67) = 81.89, p < .001$

---

1. In an extensive review of the literature regarding the Stroop effect, MacLeod (1991) confirmed that a large number of studies have replicated Stroop’s gender difference findings and provided a review of the alternative explanations. Although at that time MacLeod was not convinced these sex differences are robust, he concluded that if there are substantial differences between the sexes on Stroop task performance, it is likely due to color–naming response speed. Germane to the current study, sex-based differences have also been demonstrated in semantic and episodic memory tasks (Herlitz, Nilsson, & Backman, 1997; Kimura & Clarke, 2002; Kramer, Delis, Kaplan, O’Donnel, & Priftieta, 1997; Vonk, Eriksson, Nilsson, & Herlitz, 2003), and in naming tasks (Jensen & Rohwer, 1966; Wolff, Hurvitz, Imamura, & Lee, 1985). Given that response speed, semantic processing and naming are components of the tasks used in this study, we conducted each of the regression analyses with sex included as an independent variable. Sex did not account for unique variance in any of our analyses. For ease of readability and interpretation, we chose not to include the variable sex in the analyses reported in the results section.
(see Table 4). Again, the data indicated that after controlling for individual differences in the intercept, individual differences in the ALTM task accounted for unique variance in the improvement of the color–word task.

To ensure the relationship between improvement on the color–word task and the ALTM task was not simply due to individual differences in response latencies on tasks using response time as part of the measure of performance, we conducted a second hierarchical linear regression in which mean response time for correct responses to incongruent trials of the number Stroop task was included in block 1. If latency on the number Stroop task and performance on the ALTM accounted for the same variance in color–word task improvement then the variance accounted for by the ALTM is not likely due to facilitated performance of procedural memory, but instead by simple response time. In Table 4 we again see the full model was again significant, $F (1, 66) = 8.43, p < .001$, and that the ALTM task again accounted for unique variance in improvement in color–word performance $\Delta R^2 = .04, F (1, 66) = 8.43, p = .005$.

4. Discussion

The goal of the present study was to provide further evidence that the ALTM task effectively captures a measure of facilitation of procedural memory. Because of the hypothesized relationships, regression analyses were chosen to examine the relationships between ALTM task performance, color–word Stroop improvement across trials, and number Stroop performance. In effect this study attempted to answer two distinct questions:

1. Is the ALTM task able to effectively predict increased performance on the Stroop task across trials?
2. Does the ALTM task account for unique variance in improved performance on a cognitive operation beyond that accounted for by performance on a similar task.

Prior to beginning this study, it was hypothesized that the ALTM task would be able to predict improvement in the color–word Stroop task because it was also hypothesized that the ALTM task was, at least in part, capturing a measure of facilitation of procedural memory as was originally proposed by Woltz and Was (2006). Although Was et al. (2012) provided a clear demonstration that the ALTM task shares variance with measures of fluid intelligence and working memory capacity, no prior research used the ALTM task to predict increase in performance on a cognitively simple task in an attempt to explain what individual differences in ALTM task performance may be capturing.

The ALTM task was effective in predicting color–word Stroop improvement across trials. Although the amount of variance accounted for by the ALTM task was small, these findings are still relevant as it lends support to the idea that the ALTM task is tapping into the construct of facilitation of procedural memory. This could explain the enduring effects found by Was (2010b). While these findings cannot demonstrate any sort of causal relationship, the correlational study is still worth noting.

The ALTM associate items correlated with both congruent and incongruent response times on the color–word Stroop task in Block 2 of trials on both Days 1 and 2, as well as with congruent response times on both days. The neutral items did not correlate with any of the response times. This seems to support the argument that the mechanisms underlying performance on the associate item trials are tapping into a subject’s ability to utilize new procedures more quickly or effectively. Within the framework of the ALTM task this has been considered to be a persistent memory for prior cognitive operations, such as categorization or feature identification (Woltz & Was, 2006). In the Stroop task, such a measure should relate to a subject’s ability to adjust the selection of an appropriate, though less-often used, procedure (color-naming) relative to a more highly-preferred procedure (reading).

It was hypothesized that the associate item trials would more accurately reflect this ability specifically because they present the subject with related, but not previously seen, words in the trial and subjects are asked to make a synonym judgment. Because these items are related to previously seen words if the subject is fast and accurate at responding it may be because they can more quickly assimilate and utilize new procedural memory related to the task at hand. If this is a general ability it should relate to improvement in any number of tasks.

The Stroop task was chosen specifically because it is largely unrelated to long-term semantic priming, which is the other mechanism often postulated to explain the ALTM task’s utility (Woltz, 2010). In the Stroop task, the priming of color words might actually interfere with task performance as it may reduce the ability to inhibit the automaticity of reading activated words. Overall, the relationship between the associate item trials and larger reductions in response times across trials should support the notion that at least the associate item trials in the ALTM task are tapping into the facilitation of procedural memory.

The fact that the ALTM associate item task performance also relates to number Stroop performance could indicate one of two things. The simplest explanation is measurement of performance on both tasks incorporates response latency. However, as seen in the Results section (Table 4), number Stroop performance and improvement in color–word Stroop performance share a great deal of variance, and performance on the ALTM performance and number Stroop accounted for virtually equal amounts of variance in color–word task improvement. Furthermore, ALTM task performance accounted for unique variance in color–word improvement. Therefore, an alternative explanation is that the relationship between ALTM and number Stroop performance reflects the individual’s ability to continue to assimilate new procedural memories and apply them. It is simply another case of demonstrating the individual’s level of ability to facilitate new procedures. Clearly,

<table>
<thead>
<tr>
<th>Table 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Values for hierarchical regressions.</td>
</tr>
<tr>
<td>Predictor</td>
</tr>
<tr>
<td>---</td>
</tr>
<tr>
<td>Color–word intercept</td>
</tr>
<tr>
<td>ALTM task</td>
</tr>
<tr>
<td>$R^2$</td>
</tr>
<tr>
<td>$R^2$</td>
</tr>
<tr>
<td>$F$ for change in $R^2$</td>
</tr>
</tbody>
</table>

Note.  
* $p < .01$.  
** $p < .001$.  
### F.X. Smith Jr., C.A. Was / Learning and Individual Differences 29 (2014) 30–40
the evidence for this assumption is not conclusive and further investigation into this hypothesis is warranted. However, the findings do support the hypothesis that there is a unique relationship between improvement in simple cognitive operations and the facilitation of procedural memory as distinguished by the ALTM task.

5. Limitations

There are limitations to the present investigation. One key limitation is that there were no measures of working memory capacity or executive function (aside from the Stroop task itself) collected during the experiment. This limits the extent of the regression analysis accounting for variance in improvement on Stroop task performance. While this limits the strict interpretation of how much variance is being uniquely accounted for by ALTM task performance it does not nullify present findings. Instead, it suggests that a more thorough follow-up study may be justified in order to learn more about the variance in improvement accounted for by a number of related constructs, such as working memory, executive function, task switching, and other related cognitive constructs.

Another limitation is the use of only synonym comparisons. The ALTM task was originally developed with category comparisons and may be an important measure to include in future research as well. A more complete examination of this matter in future experiments may include structural equation modeling and use multiple types of ALTM task measures to represent facilitation of procedural memory. Using only one measure in this experiment could be viewed as a limitation.

In addition, the sample size reached in the present study is a serious limitation. Attempts at generating a latent growth curve model were unsuccessful due to the limited sample size. It is often recommended that studies using path analysis or structural equation modeling contain at least 200 participants in order to obtain accurate results (Jöreskog & Sörbom, 2006). However, the significant findings related to relationship between the ALTM associate items task and improvement on Stroop performance still warrants further investigation regardless of the small sample size used in the present study. The sample population is also a concern as only undergraduate education students participated in the present study, limiting generalizability of any conclusions drawn from this experiment.

A final consideration in terms of limitations is to recall that the present study cannot establish causality, and it would not even theoretically make sense to make such an argument. The present study sought to lend weight to the argument that the ALTM task taps some measure of facilitation of procedural memory. Performing well on the ALTM task does not cause better performance on the Stroop task. Rather, both tasks are in some way tapping the same construct.

6. Conclusion

Although the present study does not answer conclusively what the ALTM task is measuring, the findings here seem to support the notion that the ALTM task can measure an individual’s ability to acquire new procedural memory traces and effectively put them to use. Moreover, this measure of facilitation may also relate to an individual’s ability to transfer relevant knowledge to related tasks. It was stated earlier that part of the importance of identifying what the ALTM task is measuring is due to its ability to contribute to the prediction of intelligence and comprehension. Based on the present evidence, it would seem that the ALTM task might also be able to help predict more efficient skill acquisition in individuals. It may be even more effective at predicting long-term efficiency in skill acquisition. As more predictive elements are associated with the ALTM task, it becomes even more important to continue to try to understand just what constructs this task is measuring. The present study was an initial attempt to comparatively evaluate the claims that the ALTM task is measuring either facilitation of procedural memory or long-term semantic priming. The present results suggest that facilitation of procedural memory is certainly implicated in what drives performance on the ALTM task.

References


