EFFICIENT COGNITIVE OPERATIONS PREDICT SKILL ACQUISITION

A thesis submitted to the
Kent State University College
of Education, Health, and Human Services
in partial fulfillment of the requirements
for the degree of Master of Arts

by

Francis X. Smith, Jr.

August 2012
The present study examined the relationship between performance on the ALTM task and improvement in Stroop task performance across trials. Prior research indicated that the ALTM task may be capturing individual differences in facilitation of procedural memory but has often been confounded with long-term semantic priming due to the nature of the task. The Stroop task was chosen because related semantic information is largely irrelevant to performance. The study was conducted as a correlational experiment and did not utilize experimental manipulations. Path analysis revealed that ALTM task performance accounted for 11.3% of the variance in improvement in color-word Stroop improvement over two sessions. No other hypothesized relationships were significant in the path model. Results are discussed in light of the apparent relationship between ALTM task performance and an individual’s ability to acquire new procedural memory traces. Implications of these findings are discussed and future directions for continued research are proposed.
ACKNOWLEDGMENTS

I would like to offer my thanks to my committee members, Dr. John Dunlosky and Dr. Katherine Rawson, for their willingness to serve and their generous and helpful comments during the process of writing this thesis. The feedback I have received has helped to move from conceptualizing this project to completing it and for that I am grateful.

In addition I would like to thank Dr. Christopher Was, my thesis advisor, for the constant help and support he provided during this whole process. Not only has he served as my mentor throughout my progress through my degree, but he has also been indescribably helpful in my growth as a researcher as well. I am sure the lessons learned will serve me well in my continued education.

My family and friends have also given emotional and moral support throughout this whole process. I thank them for their unconditional support in this endeavor. In particular I would like to thank Jeff Harrison for his feedback throughout the project and for letting me use him as a sounding board for ideas throughout our time in our respective programs.
# TABLE OF CONTENTS

ACKNOWLEDGMENTS ........................................................................................................ iii

LIST OF TABLES ................................................................................................................. vi

CHAPTER

I. INTRODUCTION AND REVIEW OF LITERATURE ............................................... 1
   Review of the Literature ........................................................................................ 2
   Models of Working Memory ................................................................................. 4
   The ALTM Task .......................................................................................... 11
   The Stroop Task ........................................................................................... 17

II. METHOD ..................................................................................................................... 22
   Participants .......................................................................................................... 22
   Materials .............................................................................................................. 22
      Apparatus ...................................................................................................... 22
      ALTM Task ................................................................................................. 23
      Color-Word Stroop Task .............................................................................. 26
      Stroop Number Task .................................................................................... 27
   Research Design .................................................................................................. 28
   Procedure ............................................................................................................. 29

III. RESULTS ................................................................................................................... 33
   Path Analysis ....................................................................................................... 38
      Initial Model ................................................................................................. 39
      Intermediate Model ...................................................................................... 40
      Final Model .................................................................................................. 41

IV. DISCUSSION ............................................................................................................. 42
   Limitations ........................................................................................................... 45
   Conclusion ........................................................................................................... 47

APPENDICES .................................................................................................................. 49
   APPENDIX A. INITIAL PATH MODEL WITH STANDARDIZED
      PATH COEFFICIENTS ............................................................................... 50
   APPENDIX B. INTERMEDIATE PATH MODEL WITH
      STANDARDIZED PATH COEFFICIENTS .............................................. 52
APPENDIX C. FINAL PATH MODEL WITH STANDARDIZED PATH COEFFICIENTS ...............................................................................54
APPENDIX D. MAXIMUM LIKELIHOOD ESTIMATES AND FIT INDICES FOR PATH MODELS (N = 70) ..........................................................56
REFERENCES ................................................................................................................. 58
LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Mean Accuracy (in Percentages) and Response Times (in Milliseconds) for Observed Variables ((N = 70))</td>
<td>31</td>
</tr>
<tr>
<td>2. Descriptive Statistics for ALTM Task Transformations (in Correct Responses Per Minute) ((N = 70))</td>
<td>32</td>
</tr>
<tr>
<td>3. Correlations of Observed Variables ((N = 70))</td>
<td>36</td>
</tr>
</tbody>
</table>
CHAPTER I
INTRODUCTION AND REVIEW OF LITERATURE

A fundamental question in research on cognition is the identification of factors, which contribute to performance on a variety of tasks. The predictive power of any model of cognition is related to the degree to which the model can specify the cognitive components that affect performance on such tasks. Working memory measures have served as a powerful predictor of performance across such a wide variety of domains that they are often taken to closely relate to intelligence. In fact, some studies have indicated that working memory is one of the primary predictors of scores on intelligence measures (Kyllonen & Christal, 1990). However there is still variance unexplained in intelligence scores in such research. Recent work in this area has suggested that fluid intelligence, specifically, may include more than simply working memory abilities (Heitz et al., 2006). The fact that intelligence test scores can often be used to determine early tracking efforts for children in school settings suggests that it is important for both educators and researchers to understand exactly what these tests are measuring.

Some research indicates that facilitation of procedural memory accounts for unique variance in measures of fluid intelligence above and beyond that accounted for by working memory measures (Was, Dunlosky, Bailey, & Rawson, 2012). This facilitation of procedural memory represents a strengthening over time for persistent memory for prior cognitive operations. Individual differences in this ability have also been shown to relate to reading comprehension, another fundamentally important ability in school settings (Was & Woltz, 2007; Was, 2010a). The present study explores the relationship
of a measure of facilitation of procedural memory in the prediction of performance on the Stroop task across trials.

Prior work with facilitation of procedural memory conducted by Woltz and Was (2006, 2007) has primarily relied on unprimed category or feature comparisons and thus has implicitly involved available long-term memory. Performance on the Stroop task, however, is largely independent of prior knowledge or activation and thus may still reflect this facilitation of procedural memory, or memory for prior cognitive operations, without reliance on declarative long-term memory. If performance on the Stroop task increases across trials and can be accounted for by individual differences in facilitation of procedural memory, this will support the hypothesis that the ALTM task developed by Was and Woltz measures a persistent strengthening of memory for prior cognitive operations rather than some form of long-term semantic priming.

**Review of the Literature**

Numerous 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.

In an attempt to capture the contributions of related long-term memory elements after attention-driven processing within working memory, Woltz and Was (2006) developed the ALTM task. This task demonstrates the recruitment of long-term memory in performance of category, feature, and synonym comparisons of items even when the to-be-compared items were not previously processed in working memory. There is still debate about what exactly is being captured by the ALTM task. Woltz and Was originally argued that processing of exemplars from a category initiated a form of long-term semantic priming (Woltz & Was, 2006). Further examination revealed that another possible explanation could be a persistent memory for prior cognitive operations, which is strengthened during the procedures required to complete a task (Woltz & Was, 2007). Woltz has continued to pursue the role of the ALTM task in capturing long-term semantic priming (Woltz, 2010). Was, on the other hand, has investigated the ability of the ALTM task to measure a facilitation of procedural memory (Was, 2010b). The present study evaluates these two competing explanations by comparing subjects’ performance on the ALTM task to Stroop task performance.

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,
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).

The fact that 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 if the ALTM task captures a measure of facilitation of procedural memory. If the ALTM task is primarily capturing the effects of semantic priming between related long-term memory elements then there should be little, if any, correlation between performance on the ALTM task and improvement on the Stroop task. If, on the other hand, the ALTM task is capturing a facilitation of procedural memory, or persistent memory for prior cognitive operations, then performance on the ALTM task should be predictive of increased performance on the Stroop task across trials. Increased performance, in the present study, is operationalized as a decrease in latency, an increase in correct responses, or both.

Models of Working Memory

The multiple-component model and limitations. Beginning with Baddeley and Hitch’s adoption of the term in 1974, working memory has played an important role in explaining human cognition. The initial model of working memory described two separate storage components used to rehearse and maintain information within the system. The first was the phonological loop, which accounted for temporary storage of information represented phonologically. The second storage component in Baddeley’s model was the visuospatial sketchpad, which was used for visual representations of
information. The key component tying these two storage systems together was what Baddeley referred to as the central executive. As implied by the name, the central executive was responsible for coordinating a number of executive functions such as attention, processing and manipulation of information in the two stores, as well as inhibitory processes. The cognitive workspace in Baddeley’s original model of working memory was able to explain a number of important memory phenomena including results from dual-task paradigm experiments (Baddeley & Logie, 1999).

A variety of data supports Baddeley’s claim that the visuospatial sketchpad and the phonological loop are indeed separable constructs (Baddeley & Logie, 1999). Difficult mental arithmetic in the face of articulatory suppression or random letter generation demonstrates a decrease in performance while irrelevant pictures and concurrent hand movements do not produce noticeable decrements in performance (Logie, Gilhooly, & Wynn, 1994). Whereas there is significant evidence to support the structure of Baddeley’s model of working memory, the picture is clearly not complete.

The empirically demonstrated limits of the storage system’s working memory in Baddeley’s model cannot readily account for more complex cognitive procedures such as language comprehension (Anderson, 1983). Attempts to model these complex tasks in various cognitive frameworks have shown that the functional limits of working memory would need to be increased well beyond the original predictions of Baddeley’s model. In Anderson’s estimation the limits of working memory would need to be increased to 20 or more elements to simulate more complex cognitive activities within the framework of Baddeley’s original model.
Rather than directly increasing the size of working memory capacity, several theories include available long-term memory into the cognitive workspace of working memory. Baddeley himself acknowledged this shortcoming of his original model and introduced an additional component called the episodic buffer (Baddeley, 2000). The episodic buffer provides a mechanism for chunking of information as well as a means for the central executive and the two storage components to interact with relevant elements of long-term memory (Baddeley, 2003). Whereas the inclusion of such a component is sensible, there is little explanation for how the episodic buffer actually functions. Instead it is assumed to be another limited store that is controlled by the central executive and can combine multiple types of information from long-term memory into episodes to be used to effectively chunk information and allow for greater apparent temporary storage of information.

There are a number of recent models of working memory which use access to long-term memory elements to bridge the gap between empirically demonstrated limitations of working memory storage and the need for greater amounts of information to complete complex cognitive tasks (Miyake & Shah, 1999). Within the framework of this review, models, which propose access to long-term memory elements based on activation of production rules, are the focus.

The embedded processes model. It has been suggested that highly active portions of long-term memory may be more available for processing while remaining outside of the immediate focus of attention. Cowan’s model, for example, defines working memory as the activated portion of long-term memory currently in the focus of
attention (Cowan, 1999). In this representation of working memory there are highly activated portions of long-term memory that remain outside of the focus of attention yet are more available for immediate processing than less activated portions of long-term storage. This activation is time limited rather than capacity limited and thus can encompass a great deal of relevant long-term memory elements so long as the related items in working memory remain in the focus of attention (Cowan, 1999).

Within his model, Cowan introduces the concept of “virtual” short-term memory. This type of short-term memory is said to include those portions of long-term memory which are related contextually to the situation and thus can function much as if they were in an activated form regardless of actual activation level. This virtual short-term memory supports more complex cognition that would normally go beyond the limits of attention. An important feature of this access to long-term memory is that while familiar features may become active they may often do so in novel ways; this novel link between familiar features in working memory may end up being stored as its own individual long-term memory trace. Declarative memories are often assumed to require attentive processing while procedural memory may be acquired automatically assuming sufficient attention is devoted to the task itself.

For the purposes of the present study it is this virtual short-term memory that may best represent what the ALTM task captures. If a subject is focused on the task at hand, according to Cowan’s model (1999), the procedures needed to complete the task may become automatically activated and implemented without attention from the working
memory system. This should increase the amount of attention available for selective attention to task goals and inhibition of any irrelevant responses.

It is worth noting that Oberauer (2002) held a very similar view of working memory referred to as a concentric model. In this model working memory consists of activated long-term memory elements, a capacity-limited region of direct access, and the focus of attention. The model is considered a concentric model because each level lies within the region preceding it. The focus of attention lies within the region of direct access, which in turn lies within the content of activated long-term memory.

**Long-term working memory.** An alternative theory of long-term memory’s support of functions within working memory is presented by Ericsson and Kintsch (1995). Rather than proposing an activation-based model of long-term memory’s support, Ericsson and Kintsch proposed a contextual and skill-based use of storage known as long-term working memory. The premise for inclusion of long-term working memory is that traditional short-term working memory models fail to take into account the ability of an individual to rapidly encode information in long-term memory during skilled performance in order to use said information at a later time with appropriate cues, effectively increasing working memory capacity (Ericsson & Delaney, 1999).

Long-term working memory allows for durable storage of relevant information, which can later be efficiently accessed when the appropriate cues are present in short-term working memory (Ericsson & Kintsch, 1995). It is because of this cue-based approach that the model is contextual and primarily explains performance in routine scenarios. If the context of task performance varies, the cues in short-term memory may
not reliably provide access to the relevant information from long-term working memory. Another noteworthy feature of long-term working memory is that with such expert performance it becomes difficult to make the traditional distinction between memory, knowledge, and procedures as they are all effectively grouped together (Ericsson & Delaney, 1999).

Within the present study the concept of long-term working memory may also relate to performance on the ALTM task. Whereas Ericsson and Kintsch (1995) primarily addressed expert performance in their presentation of long-term working memory, it is likely that repeated performance of a task, such as the Stroop task, within similar settings may lead to the recruitment of long-term working memory to facilitate better performance. Although the Stroop task does not demand a large cognitive load in terms of individual items to be maintained in working memory, it places high demands on selective attention and inhibition and could be aided by the support of long-term working memory.

**The ACT-R framework.** While not strictly speaking a model of working memory, Anderson’s ACT-R model was developed to explain human performance during cognitive activities that could not easily 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 nodes that are
more highly activated and can thus be easily accessed relative to other less activated declarative nodes. Both declarative 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. Considering repeated performance on the Stroop task, the ACT-R framework has already been used to model performance on the Stroop, which allows for comparison of the present study to the simulations from ACT-R (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. If the ALTM task is capturing a measure of facilitation of procedural memory it should reflect this strengthening as proposed by Altmann and Davidson.

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. 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. 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 the models discussed above (Anderson, 1983; Cowan, 1999; Ericsson & Kintsch, 1995).

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 memory load phase of the task can affect priming 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 priming 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 (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 priming 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 priming (Woltz & Was, 2006). Subjects asked to remember a category show greater priming for later exemplars related to that category.
Interestingly, subjects asked to ignore a category also show greater priming 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 priming effects rather than simply the maintenance of the exemplars. In order 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, priming effects have been shown to be greater for the ignored category (Woltz & Was, 2006).

Recall of the appropriate memory load words has been included in many previous ALTM task experiments primarily to validate the assumption that subjects have maintained the intended exemplars for the duration of the task between presentation and recall (Woltz & Was, 2006). Across all prior research with the ALTM task utilizing this recall phase subjects have demonstrated relative accuracy in recalling the exemplars they were instructed to remember (Was, 2010b; Woltz & Was, 2006, 2007). An alternative to the recall phase has also been used. In this alternative design subjects were not instructed to remember or ignore either category and were instead simply presented with the memory load. Following memory load presentation subjects were asked which concept had been presented with the most instances. This alternative was presumed to require subjects to perform category identification for both categories in the list instead of just one category of instruction (Was & Woltz, 2007).

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 have 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).

Numerous manipulations of the general experimental technique have been described above to evaluate what the task itself may be measuring. During experiments in which subjects were instructed to remember one set of category exemplars, even the ignored category showed significant priming effects, which suggests that even minimal attention-driven processing in working memory produces priming of semantically related concepts (Woltz & Was, 2006). When subjects are shown category exemplars but later presented category features for comparison, however, there are no significant priming effects (Woltz & Was, 2007). These results led to the conclusion by Woltz and Was (2007) that the ALTM task taps a measure of persistent memory for prior cognitive
operations. If this view is accurate, then it is procedural memory rather than semantic content that is being strengthened and demonstrated in the ALTM task.

In addition to the effects of presenting either categories or features during the memory load and requiring comparisons of a different type of related information, the duration of the priming effects found in ALTM tasks indicates that it is more likely the facilitation of procedural memory that is being measured. Was (2010b) found that priming effects could still be found even when the comparison phase was presented 24 hours after the initial memory load phase of the task. Typical accounts of long-term semantic priming cannot account for these durations, although facilitation of procedural memory is typically expected to last longer (Was, 2010b).

**Uses of the ALTM task.** Prior research has demonstrated that the ALTM task likely measures a form of facilitation of procedural memory (Was, 2010b; Woltz & Was, 2007). More recent research utilizing the ALTM 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. Whereas 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 represents 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 subjects’ facilitation of procedural memory then the ALTM task and the Stroop task
performance should be positively related. As explained by the ACT-R model, in the context of the present study the ALTM task may be measuring the degree to which a subject can adjust the utility value of color naming compared to reading text for the purposes of the task (Altmann & Davidson, 2001).

Although there is evidence that the ALTM task may be measuring a form of facilitation of procedural memory, the results are far from conclusive. A recent experiment by Woltz (2010) separated semantic priming and repetition priming and found that semantic priming may reflect contributions solely from a strengthening of abstract semantic representations while repetition priming likely involves some non-declarative memory for prior cognitive operations. Furthermore Woltz suggested that both short- and long-term semantic priming effects may be able to be represented by the same mechanism. Semantically related information is certainly important to what is being captured by the ALTM task, whether it is strictly long-term semantic priming or also a facilitation of procedural memory. This study used the Stroop task measures obtained to evaluate whether or not ALTM is effectively capturing a strengthening of memory for prior cognitive operations.

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). In the study by Was, the ALTM task’s contribution to reading comprehension variance is explained as an increased availability and better access to background knowledge which can support
reading comprehension. 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 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 et al., 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.

**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 regards
to the present study. First, there are 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). Second, 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. It may also account for a more efficient inhibition of the prepotent response of word reading.

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 in his initial study. 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 neglect should certainly play some role in overall Stroop performance, it is likely an all-or-nothing effect. Either a subject will successfully
maintain the goal and use it to direct activity or they will not maintain the goal and choose an inappropriate response. This does not adequately represent the interference effects of the Stroop task, even if it can be used to predict some aspects of the behavioral data available. 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 (color naming) and the automatic process (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.

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. While 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.
CHAPTER II

METHOD

Participants

Participants in this study were recruited from the undergraduate student population of the College of Education, Health, and Human Services at Kent State University. Recruitment took place in Educational Psychology courses (EPSY 29525) during the Fall 2011 and Spring 2012 semesters. Students who chose to participate in the study were instructed in the use of the Educational Psychology Research Participation Sign-up System and asked to register in order to attend sessions in the Educational Psychology Laboratory. All participants who completed both required sessions received course credit as compensation for their time. Participants all had either normal or corrected to normal vision.

A total of 82 undergraduate students (54 females, 23 males, 5 missing demographic information) 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.

Materials

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.

**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 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 priming 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 were 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. The five words in the memory load were presented for 1500 ms each, with a 500 ms blank screen and an attention priming 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] or more words that meant [Concept2]?” 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 priming (an actual item from the memory load) in both positive and negative matches, as well as an indirect (related item not presented in the memory load) positive match. The set also included the less numerous concept in the form of direct priming positive and negative matches, as well as an indirect positive
match. The final two items were from a neutral concept (not presented in the memory load at all) positive and negative match. 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 experiment 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 effects of indirect priming of related concepts to the memory load compared to response speed and accuracy of neutral items not previously presented during the memory load.

**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. Subjects were given 16 warm up trials in which they were instructed to respond to words printed on the screen. 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. They were given 32 warm up trials to practice their responses before proceeding to data collection trials. In both cases the presentation was identical. There was a 500 ms wait between each trial, followed by an attention stimulus consisting of an asterisk displayed in the location of the word to be displayed. This asterisk was shown for 500 ms and was immediately followed by the word printed in the appropriate ink. Each block of trials consisted of 32 trials, 24 of which were congruent trials and 8 of which were incongruent. The available ink colors (and word presentations) were “RED,” “YELLOW,” “GREEN,” and “BLUE.” Each subject completed two sets of three blocks for a total of 192 trials, 75% of which were congruent. There was a break between sets which was self-paced. After each individual response a subject was informed if their response was correct or incorrect. If correct, their response speed was displayed. If incorrect both response speed and the appropriate response were displayed.

**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 to place four fingers of one hand on the number keys 1-4 at the top of the
keyboard. Only the numbers 1-4 were used as stimuli, and they were presented in
random order within blocks. Instructions emphasized response speed while minimizing
errors as in the color-word Stroop task. Following correct responses, latency feedback
was provided for one second. After incorrect responses, the word “incorrect” was
presented for one second. Average latency was provided at the end of each block.

Part two was similar in format, except that character strings from one to four
characters in length were presented and participants were instructed to respond with the
number of characters presented not the value of the characters. For each of four string
lengths, there were five possible characters: 1, 2, 3, 4, and X. All characters within a
string were the same (e.g., ‘33’, ‘XXXX’, ‘111’, ‘22’, etc.).

There were four blocks of 20 trials each in part two. Three different trial types
corresponded to those in the traditional Stroop task. Of the 20 trials in each block, 12 had
content designed to interfere with the length judgment (e.g., ‘2’, ‘3’, ‘4’, ‘11’, ‘33’, ‘44’,
to facilitate the length judgment (i.e., ‘1’, ‘22’, ‘333’, and ‘4444’). Finally, four trials
contained content that was neutral with respect to length judgment (i.e., X, XX, XXX,
and XXXX). Trial format and feedback were the same as described in part one.

**Research Design**

This study was designed to examine the relationship between performance on the
ALTM task, which is hypothesized to measure facilitation of procedural memory, and
improvement in performance across trials on the Stroop task. Sex was also included in the analysis to examine possible differences between males and females. In the present study, individual differences in ALTM task performance were hypothesized to relate to the degree of improvement in Stroop task performance. Because of the nature of this study, correlational design was deemed appropriate and neither control groups nor random assignment were employed.

**Procedure**

Participants in the study completed the above tasks in two one-hour sessions in the educational psychology laboratory. The second session occurred approximately 24 hours 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.

All data were entered into the Statistics Package for the Social Sciences (SPSS version 18.0) for analysis. The LISREL 8.80 (Jöreskog & Sörbom, 2006) software package was used to complete the path analysis portion of the study. Both correlation and path analysis were used to examine the relationship between the variables included. Prior to evaluating any model for path analysis, descriptive statistics were analyzed and
assumptions of normality for each observed variable were checked. 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 priming 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 were indicated to be non-normal in distribution ($p < .05$). The Shapiro-Wilk 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: 

$$\text{Accuracy} \left( \frac{\text{ResponseTime}}{60000} \right)$$

Table 1
### Mean Accuracy (in Percentages) and Response Times (in Milliseconds) for Observed Variables ($N = 70$)

<table>
<thead>
<tr>
<th>Variables</th>
<th>Accuracy</th>
<th>Response Time</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$M$</td>
<td>$SD$</td>
</tr>
<tr>
<td><strong>Color Stroop</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Day 1 Congruent</td>
<td>96</td>
<td>4</td>
</tr>
<tr>
<td>Day 1 Incongruent</td>
<td>84</td>
<td>11</td>
</tr>
<tr>
<td>Day 2 Congruent</td>
<td>95</td>
<td>4</td>
</tr>
<tr>
<td>Day 2 Incongruent</td>
<td>85</td>
<td>13</td>
</tr>
<tr>
<td><strong>ALTM Task</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Memory Load Items</td>
<td>88</td>
<td>11</td>
</tr>
<tr>
<td>Associated Items</td>
<td>89</td>
<td>9</td>
</tr>
<tr>
<td>Neutral Items</td>
<td>81</td>
<td>10</td>
</tr>
<tr>
<td><strong>Number Stroop</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Congruent</td>
<td>98</td>
<td>3</td>
</tr>
<tr>
<td>Incongruent</td>
<td>88</td>
<td>10</td>
</tr>
</tbody>
</table>
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>

Both the Kolmogorov-Smirnov and the Shapiro-Wilk 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).

It was determined that a number of different fit indices should be used, as recommended by Schumacker and Lomax (2010) in their text on structural equation modeling. Five indices were chosen to evaluate the path model including the Chi-Square (\(\chi^2\)), Goodness of Fit Index (GFI), Adjusted Goodness of Fit Index (AGFI), Root-Mean-Square Error of Approximation (RMSEA), and the Standardized Root-Mean Residual (SRMR). It is common to use more than one index to evaluate a model as there is no consensus that one index will always be best for such an evaluation.
CHAPTER III

RESULTS

Prior to any analysis of relationships between the proposed variables, it was important to verify that expected priming occurred for both the ALTM task and the Stroop tasks implemented in the current study. Dependent-samples t-tests were used to confirm that there were differences in performance across Color-Stroop trials, differences between congruent and incongruent trial performance, and also between associated items and neutral items on the ALTM task.

A dependent-samples t-test on Stroop performance on day one between congruent response times ($M = 662, SD = 88$) and incongruent response times ($M = 861, SD = 145$) revealed a significant difference between response time performance for the conditions at the specified .05 level, $t(79) = -19.21, p < .001, d = -1.66, 95\% CI [-219, -178]$. Similarly, a dependent-samples t-test on Stroop performance on day two between congruent response times ($M = 611, SD = 78$) and incongruent response times ($M = 761, SD = 118$) also revealed a significant difference in performance at the .05 level, $t(73) = -16.44, p < .001, d = -1.50, 95\% CI [-169, -132]$. These results indicate that there is a significant Stroop effect occurring in that there is a significant difference between response time performance on congruent and incongruent responses. These results are as expected in the color-word Stroop task.

Another dependent-samples t-test was conducted to examine the difference in performance on incongruent trials between days one and two. The test revealed that incongruent trial response time on day one ($M = 848, SD = 140$) and incongruent trial
response time on day two ($M = 761, SD = 119$) were significantly different at the specified .05 level, $t(70) = 8.88, p < .001, d = .67, 95\% \text{ CI}[68, 107]$. There was a noticeable improvement on incongruent trial response time performance between the two days. However, to avoid concerns about re-learning occurring at the beginning of day two due to the time interval between sessions, it was decided to utilize color-word Stroop performance from day one block one and compare that to day two block two performance. This should eliminate the interference from subjects re-familiarizing themselves with the task upon returning on day two.

The final dependent-samples $t$-test on color-word Stroop performance revealed that day one block one incongruent response time ($M = 996, SD = 205$) and day two block two incongruent response time ($M = 738, SD = 121$) were indeed significantly different and the specified .05 level as expected, $t(70) = 12.17, p < .001, d = 1.53, 95\% \text{ CI}[215, 300]$. Based on the difference in results it is apparent that there were some re-learning effects when subjects returned on day two to complete the task a second time. It was decided to use the difference in day one block one performance compared to day two block two to operationalize Stroop task improvement across trials. Rather than operationalizing Stroop performance as simple difference scores, however, it was decided that a linear regression would be used to predict day two block two performance on incongruent trials based on day one block one performance. Residual scores were generated and saved for each subject during the regression calculation and these residuals were used instead of simple difference scores.
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 was a difference between performance on associated items and neutral items, another dependent samples $t$-test was performed. The test revealed that the transformation for associated items ($M = 46.6, SD = 12.9$) and the transformation for neutral items ($M = 34.2, SD = 8.7$) were significantly different from one another, $t(78) = 13.63, p < .001, d = 1.13, 95\% \text{ CI} [10.6, 14.3]$. 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.

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, improvement in color-word Stroop performance across trials, and generalization to number Stroop performance. The results are presented in Table 3.
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>
<th>10.</th>
<th>11.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. ALTM Memory</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. ALTM Associate</td>
<td>.85**</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. ALTM Neutral</td>
<td>.85**</td>
<td>.79**</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. CS Day1 Block1 Incongruent RT</td>
<td>-.12</td>
<td>-.16</td>
<td>-.05</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. CS Day1 Block1 Congruent RT</td>
<td>-.24*</td>
<td>-.25*</td>
<td>-.08</td>
<td>.60**</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. CS Day2 Block2 Incongruent RT</td>
<td>-.11</td>
<td>-.27*</td>
<td>-.04</td>
<td>.50**</td>
<td>.59**</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. CS Day2 Block2 Congruent RT</td>
<td>-.17</td>
<td>-.27*</td>
<td>-.09</td>
<td>.34**</td>
<td>.69**</td>
<td>.75**</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8. NS Incongruent RT</td>
<td>-.24*</td>
<td>-.38**</td>
<td>-.22</td>
<td>.23</td>
<td>.41**</td>
<td>.43**</td>
<td>.52**</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9. Residual Scores from CS Regression</td>
<td>-.06</td>
<td>-.22</td>
<td>-.02</td>
<td>-.01</td>
<td>.33**</td>
<td>.86**</td>
<td>.67**</td>
<td>.36**</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10. Residual Scores from NS Regression</td>
<td>-.09</td>
<td>-.22</td>
<td>-.09</td>
<td>.09</td>
<td>.06</td>
<td>.24*</td>
<td>.15</td>
<td>.56**</td>
<td>.23</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>11. Sex</td>
<td>-.02</td>
<td>-.09</td>
<td>-.08</td>
<td>.18</td>
<td>.02</td>
<td>.13</td>
<td>.02</td>
<td>.07</td>
<td>.05</td>
<td>.13</td>
<td>1</td>
</tr>
</tbody>
</table>

*Note.* * = \( p < .05 \), ** = \( p < .01 \)

Performance on neutral items on the ALTM task did not correlate 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 correlated with day two block two
response times and also day one block one congruent trials but not with incongruent trials. This also would be expected if associated items are effectively capturing a measure of facilitation of procedural memory. Associated item performance should be more related to improvement than initial performance. Interestingly, associated item performance also correlated with number Stroop incongruent response times. This could lend support to the concept that individual differences in facilitation of procedural memory may allow for more rapid generalizability of Stroop performance. Finally, it is worth noting that the memory load item performance on the ALTM task did not correlate with either improvement on the Stroop task (based on residual scores) or later performance, though it did correlate with initial performance on congruent trials.

Because it seemed that performance on the associated items trials was related to improvement in performance whereas performance on neutral items was not, a linear regression was carried out using neutral item performance to predict associate item performance. Just as with the color-word Stroop regression, individual standardized residual scores were saved for each subject. These residuals operationalize the measure of facilitation of procedural memory in the current study because they essentially measure performance on associated items while controlling for performance on neutral items. The same technique was used to generate a standardized residual score for response times on the number Stroop task. Congruent performance was used to predict incongruent performance, thus essentially controlling for a subject’s overall speed in an attempt to only examine ability to generalize selection instructions from one Stroop task (color-word) to another (number). Based on the correlational analysis, a hypothetical
Path model (see Appendix A) was tested to examine the relationships between several of these variables.

**Path Analysis**

A total of three models were run to analyze the relationship between sex, ALTM task performance, increase in color-word Stroop performance, and generalizability to number Stroop. To form the initial model, it was hypothesized that ALTM task performance should relate to an increase in color-word Stroop task performance. It was also hypothesized that sex may relate to this increase in color-word Stroop task performance, and possibly to ALTM task performance. Finally, number Stroop performance was only hypothesized to be predicted by the improvement made in color-word Stroop task performance.

Further modifications to the path model were made due to sex not showing any significant relation to any other observed variable. However, given the theoretical support for including sex and the presence of a fairly large standardized residual, it was decided that a path from ALTM task to number Stroop task performance be added prior to eliminating sex. Only one modification is recommended at a time so this resulted in an intermediate model. After analysis it was determined that, in order to make the model as parsimonious as possible, sex should indeed be eliminated. Unfortunately the final model was saturated because of the exclusion of sex and thus could not be evaluated by fit indices. See Appendix D for a comparison of the three models.
Initial Model

The initial model only had one significant path ($p < .05$), which was the path from the ALTM task residual to color-word Stroop residual. The path from the color-word Stroop residual to the number Stroop residual approached significance and would have achieved it had the entire sample size been included in the analysis. The path from sex to either ALTM task residual or color-word Stroop residual was non-significant ($p > .05$) and the standardized effects of sex on either of these variables were almost non-existent. Appendix A contains the initial model including standardized path coefficients.

Despite the lack of more significant paths, the model was able to account for 11.4% of the variance in increase in color-word Stroop performance over time, though it only accounted for 5.1% of the variance in number Stroop performance. In addition, the Normal Theory Weighted Least Squares Chi-Square ($\chi^2$) test indicated that there was not a significant difference between the sample covariance matrix ($S$) and the model’s population covariance matrix ($\Sigma$), ($\chi^2 = 3.43$, $df = 2$, $p = .18$). As mentioned above, however, there is no one perfect way to measure model fit and there were several non-significant paths. In addition there was a fairly large standardized residual (-1.56) between ALTM task performance and number Stroop task performance, indicating a potential relationship that was not being accounted for.

Before proceeding with further analysis, it was decided to add a path from ALTM task performance to number Stroop performance. Theoretically this decision was supported because if ALTM is capturing a measure of facilitation of procedural memory then it should relate to an individual’s ability to transfer procedures between different
Stroop tasks as well as more generally allowing a subject to increase performance more quickly (such as within one trial block).

**Intermediate Model**

In the intermediate model, a path from ALTM task performance to number Stroop performance was added. This was added prior to deleting sex partially due to theoretical considerations but also because adding this path without sex would lead to a saturated model and prevent comparisons. See Appendix B for the intermediate model including standardized path coefficients. There was still only one significant path from ALTM task to color-word Stroop task \((p < .05)\); however both the path from color-word Stroop to number Stroop as well as the path from ALTM task to number Stroop approached significance even with the small sample size in the present study. As can be seen from comparing the models in Appendices A and B, the addition of a path from ALTM task to number Stroop also somewhat attenuated the effects of color-word Stroop on number Stroop. Thus suggests that the ALTM task measure really may underlie both the increase in color-word Stroop performance and also either transfer of that procedural memory to a related task or more immediate facilitation in a new task.

Similarly to the previous model, the intermediate model accounted for 11.4% of the variance in color-word Stroop improvement across trials and also accounted for 8.6% of the variance in number Stroop performance. This suggests a role for the ALTM task in predicting skill acquisition and possibly in encouraging transfer of training between related tasks. Again, the Normal Theory Weighted Least Squares Chi-Square \((\chi^2)\) test indicated that there was not a significant difference between the sample covariance
matrix ($S$) and the model’s population covariance matrix ($\Sigma$), ($\chi^2 = .945, df = 1, p = .33$).

The RMSEA value for the model was 0.0, which is below the usual limit of .05 (Schumacker & Lomax, 2010). The SRMR was also below the recommended .05 with a value of .04, though that puts it close to the upwards acceptable limit. The GFI for the model was .993, which puts it above the criteria of .95 (Schumacker & Lomax, 2010). However the AGFI was .932, which is a bit below the commonly accepted value.

Still, four out of five measures indicate the model fits the sample data well. However path models, as with any theory, should remain as parsimonious as possible while still explaining as much variance as possible. Given that the paths from sex to ALTM task performance and color-word Stroop improvement were not only non-significant but also seemed to explain only a very small portion of the variance, it was decided that sex should be eliminated from the current model. This leaves a saturated model which is presented briefly to examine what changes there were in variance accounted for by the model.

**Final Model**

Because the final model was saturated it was a perfect fit for the data with zero degrees of freedom. It is worth noting, however, that even when sex was excluded from the model 11.3% of the variance in color-word Stroop improvement was still accounted for as well as 8.6% of the variance in number Stroop performance. Eliminating one out of the four observed variables in the model reduced the variance accounted for by 0.1%. See Appendix C for the final model with standardized path coefficients.
CHAPTER IV
DISCUSSION

The goal of the present study was to identify whether or not the ALTM task was effectively capturing a measure of facilitation of procedural memory. Because of the hypothesized relationships, path analysis was chosen to examine the relationships between sex, ALTM task performance, color-word Stroop improvement across trials, and number Stroop performance. In effect this study attempted to answer three distinct questions:

1. Is the ALTM task able to effectively predict increased performance on the Stroop task across trials?
2. Does practice on one type of Stroop task transfer to another version of the task?
3. Are there sex-based differences in either of the two above effects?

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). No prior research used the ALTM task to predict increase in performance on a cognitively simple task in an attempt to explain what the ALTM task is actually tapping into.

It was also hypothesized based on prior research that training on the color-word Stroop task might transfer to another version of the task such as the number Stroop task (Davidson et al., 2003). If this training and transfer did occur it was of interest in the
present study how much of that transfer might be related to ALTM task performance. Although the present study cannot conclusively answer to what degree the ALTM task underlies potential transfer, it is clear that the relationship between color-word and number Stroop performance is in some way affected by the ALTM task performance as seen in the intermediate and final models.

The final question was of interest primarily because of the original explanation for part of the Stroop effect (Stroop, 1935). In his original research, Stroop hypothesized that part of the increase in color-naming ability was related to sex of the subject. It was thought that, if this were really the case, perhaps it was not strictly color-naming that is influenced by sex but possibly more generally a utility value adjustment ability as presented in the ACT-R simulation of the Stroop task (Altmann & Davidson, 2001).

Whereas sex was not significant in any of the hypothesized relationships, it was still found that the ALTM task was effective in predicting color-word Stroop improvement across trials. Even though the path model itself ended up being reduced to a very simple form, 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). Regardless of the simplicity of the path model, the correlational study is still worth noting.

The ALTM associate items correlate with both congruent and incongruent response times on the color-word Stroop task on day two, but only with congruent response times on day one. Beyond that, the neutral items do not correlate with any of the response times. This seems to support the argument that there is something about
associate item trials which is 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 inhibit a prepotent response (reading) in favor of a more difficult and slower response (color-naming).

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 traces 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. Most likely it simply 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. However, the fact that the relationship between color-word Stroop performance and number Stroop performance was attenuated when the path from ALTM task to number Stroop task was added could indicate that facilitation of procedural memory may also underlie aspects of transfer. This was not a specific research question when organizing this study but it is certainly worthy of future attention.

Limitations

There are limitations to the present experiment. 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 path analysis in accounting for variance in improvement on Stroop task performance. It also violates one of the assumptions of path analysis which is that all relevant variables are included in the model. 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 that may be an important measure to include in future research as well. The attribute ALTM task may also be relevant in future work. In the attribute ALTM task subjects identify whether or not presented words represent attributes of similar exemplars or not. A more complete examination of this matter in future experiments may include structural equation modeling and use all three 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. A path analysis for this small of a sample size is not handled regularly within the LISREL software package and it is often recommended that studies of this nature 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. If anything, a larger sample would have likely lead to the paths from ALTM task to number Stroop and from color-word Stroop to number Stroop to be significant. 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. In the future, it may be worth considering utilizing latent growth curve modeling to better observe improvement over time.

**Conclusion**

Although the present study cannot possibly 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 may also be able to help predict more efficient skill acquisition in individuals. The associate items trials in the ALTM task, even after controlling for performance on neutral items, were able to predict 11.3% of the variance in improvement on the color-word Stroop task during less than 400 trials over the period of two days. 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 competitively 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.
APPENDICES
APPENDIX A

INITIAL PATH MODEL WITH STANDARDIZED PATH COEFFICIENTS
Appendix A

Initial Path Model With Standardized Path Coefficients
APPENDIX B

INTERMEDIATE PATH MODEL WITH STANDARDIZED PATH COEFFICIENTS
Appendix B

Intermediate Path Model With Standardized Path Coefficients
APPENDIX C

FINAL PATH MODEL WITH STANDARDIZED PATH COEFFICIENTS
Appendix C

Final Path Model With Standardized Path Coefficients
APPENDIX D

MAXIMUM LIKELIHOOD ESTIMATES AND FIT INDICES FOR PATH MODELS ($N = 70$)
### Appendix D

**Maximum Likelihood Estimates and Fit Indices for Path Models (N = 70)**

<table>
<thead>
<tr>
<th>Description</th>
<th>Initial Model</th>
<th>Intermediate Model</th>
<th>Final Model</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Paths</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sex → Color-Word Stroop</td>
<td>.03</td>
<td>.03</td>
<td>--</td>
</tr>
<tr>
<td>ALTM Task → Color-Word Stroop</td>
<td>-.34</td>
<td>-.34</td>
<td>-.34</td>
</tr>
<tr>
<td>ALTM Task → Number Stroop</td>
<td>--</td>
<td>-.20</td>
<td>-.20</td>
</tr>
<tr>
<td>Color-Word Stroop → Number Stroop</td>
<td>.23</td>
<td>.16</td>
<td>.16</td>
</tr>
<tr>
<td><strong>Equation Error Variances</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Color-Word Stroop</td>
<td>.89</td>
<td>.89</td>
<td>.89</td>
</tr>
<tr>
<td>Number Stroop</td>
<td>.95</td>
<td>.91</td>
<td>.91</td>
</tr>
<tr>
<td><strong>Independent Variables</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sex</td>
<td>.21</td>
<td>.21</td>
<td>--</td>
</tr>
<tr>
<td>ALTM Task</td>
<td>1.03</td>
<td>1.03</td>
<td>1.03</td>
</tr>
<tr>
<td><strong>Fit Indices</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\chi^2$</td>
<td>3.43</td>
<td>.945</td>
<td>0.0</td>
</tr>
<tr>
<td>RMSEA</td>
<td>.103</td>
<td>0.0</td>
<td>--</td>
</tr>
<tr>
<td>SRMR</td>
<td>.066</td>
<td>.035</td>
<td>--</td>
</tr>
<tr>
<td>GFI</td>
<td>.976</td>
<td>.993</td>
<td>--</td>
</tr>
<tr>
<td>AGFI</td>
<td>.879</td>
<td>.932</td>
<td>--</td>
</tr>
</tbody>
</table>
REFERENCES
REFERENCES


*Neuropsychology and Cognition, 10*(2), 85–98.


