EFFECTS OF STIMULUS CLASS ON SHORT-TERM MEMORY WORKLOAD IN COMPLEX INFORMATION DISPLAY FORMATS

by

Kay Chuan Tan

Dissertation submitted to the Faculty of the
Virginia Polytechnic Institute and State University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Industrial Engineering and Operations Research

APPROVED:

Robert J. Beaton, Chairperson

Helen J. Crawford

Harry L. Snyder

Beverly H. Williges

Robert C. Williges

May, 1990

Blacksburg, Virginia USA

EFFECTS OF STIMULUS CLASS ON SHORT-TERM MEMORY WORKLOAD IN COMPLEX INFORMATION DISPLAY FORMATS

by

Kay Chuan Tan

Robert J. Beaton, Chairperson

(ABSTRACT)

The objective of this research effort was to identify opportunities and demonstrate methods to reduce aircraft crew member cognitive workload (CWL) by reducing short-term memory (STM) demand. Two experiments qualitatively and quantitatively compared memory loading as a function of *stimulus class*. Experiment 1 employed a dual-task paradigm where the primary task was compensatory tracking used to load STM and the secondary task was item recognition using the Sternberg paradigm. Experiment 2 employed a single-task paradigm using a modified version of the Sternberg task. Digits, letters, colors, words, and geometrical shapes were tested as memory-set (MSET) items in the Sternberg task. Recognition latency and error rate served as objective measures of STM performance while the Subjective Workload Assessment Technique (SWAT) was employed as a subjective second measure. Root Mean Square error was used to gauge tracking performance.

Analyses of the experiments' results revealed that recognition latency and SWAT ratings statistically varied as functions of stimulus class, MSET size, and the interaction between stimulus class and MSET size. Error rate was not statistically different across stimulus class or MSET size. Post-hoc analyses found SWAT to be a more sensitive STM measurement instrument than recognition latency or error rate. No statistically significant degree of secondary task intrusion on the tracking task was found.

In addition to the commonly used classes of digits and letters, this research demonstrated that colors, words, and geometrical shapes could also be utilized as MSET items in short-term memory workload investigations. This research has, more importantly, provided further support for the vital link between STM demand and perceived workload.

The main conclusion of this research is that stimulus class optimization can be a feasible method for reducing STM demand. Differences in processing rate among stimulus classes are large enough to impact visual display design. For many context-specific applications, it should be possible to determine the most efficient stimulus class in which to portray the needed information. The findings of this research are especially applicable in situations of elevated STM demand (e.g., aviation systems operations). In general, however, the results provide helpful information for visual display designers.

ACKNOWLEDGEMENT

This research was funded by the USAF Armstrong Aerospace Medical Research Laboratory, Dayton, Ohio (Contract No. F33615-88--D-0532). Mr. Gilbert Kuperman and Ms. Denise Wilson served as technical monitors. Dr. Robert Beaton was the principal investigator.

Dr. Helen Crawford, Dr. Harry Snyder, Mrs. Beverly Williges, and Dr. Robert Williges are gratefully acknowledged for their insights and contributions to this research. The author acknowledges also the support of his co-research assistant, Dr. Michael Kahn, throughout the entire project.

The author is indebted to Dr. Beaton for his support, encouragement, and guidance as an advisor and a friend. Dr. Beaton's untiring dedication to his students and his work served as this author's inspiration to carry through the sometimes trying phases of this research effort.

This dissertation is dedicated to the author's family, especially his mother and father, wife Virginia, and son Crystan.

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

LIST OF TABLES

INTRODUCTION

Recent decades have witnessed tremendous technological strides in the aviation community. With increasingly sophisticated flight systems technologies, the performance envelopes of the aircraft as well as the crew members who fly them are being pushed to and beyond what was thought unattainable only a few years ago. Given crew members' behavioral and cognitive limitations, coupled with an acute emphasis on flight safety, it is not surprising that human factors engineering plays a crucial role in aviation systems design and operation. Human factors engineering involvement in various aviation research and development issues include the following (Wiener and Nagel, 1988):

1. Systems safety at the cockpit crew member interface;
2. Flight performance and interaction among crew members;
3. Human error in aviation systems operation;
4. Software interface and crew station design;
5. Cockpit and flight automation; and
6. Crew member cognitive workload.

The above areas of concern are linked together integrally. For example, cockpit automation cannot be implemented without assessing the impact on crew member workload. Also, the safety of crew members and their ability to interrupt and regain flight control in case of equipment malfunction are critical. Interface design issues must also be considered for the displays and controls in the cockpit.

Given the continuing advances of electronic technologies, higher degrees of automation and expert assistance will prevail and expand in aircraft applications. Thus, crew members increasingly will be relegated to roles as flight programmers, monitors, decision makers, and systems managers. Such role transformations from manual piloting to flight information processing places increasing demands on the higher-order cognitive processes. Two research areas in human performance have particular relevance in this regard: human memory and cognitive workload. The current research work focuses on practical display system design stemming from these two scientific areas.

Focus of this Research

Human memory mechanisms are important to crew station design since operators are limited capacity and bandwidth information processors. In particular, dynamic flight and status information is stored, cognitively evaluated, and retrieved in short-term memory. Excessive cognitive workload may result from information processing demands that exceed short-term memory capacity and, hence, degrade crew members' task efficiency (Kuperman and Wilson, 1986).

The framework of this research involves aircraft crew member cognitive workload (CWL) associated with tasks involving short-term memory (STM) processes. The aim is to identify and demonstrate opportunities to reduce crew member workload by reducing STM demand. As a prior experiment related to the current work, Reinhart, Glynn, Dye, Takahama, and Snyder (1988) provided experimental groundwork and validation for the use of STM concepts in the design of workload-reducing crew station information management technologies. They provided the following list of viable strategies for reducing STM workload demand:

1. Grouping and the hierarchical organization of information classes;
2. Elimination of distractor tasks and the use of redundant visual cuing;
3. Release from proactive interference;
4. Use of rehearsal strategies;

5. Multiple input, process, and output channels; and
6. Optimization of stimulus class.

Specific issues to be addressed. This research focuses on the optimization of stimulus class attributes (i.e., Strategy 6). Specifically, the feasibility of using stimulus class (i.e., digits, letters, colors, words, and shapes) as an information portrayal variable for reducing STM workload demands was examined. To elaborate, psychologists have demonstrated that the STM concepts of scanning time and memory span vary as functions of stimulus class (Cavanagh, 1972; Sternberg, 1975). For example, digits are stored in STM easily and, thus, humans have a larger memory span for this stimulus class than for shapes. The time needed to scan a digit also is shorter than the time to scan a shape. Given the variety of ways to present information by manipulating stimulus class, this research addressed the following issues. First, what numerical differences in scanning times and memory spans are associated with various stimulus classes? Second, what quantitative differences do these memory loadings incur in terms of cognitive workload? Third, what STM-related constraints and limitations are involved in using stimulus class as an information optimization technique?

The three issues above embody the general concern of *differential processing efficiency*. The ultimate worth of this research would be measured through a set of display formatting recommendations that can be generated as a result. However, note that recommendations can be justified in economic terms only if the differential processing efficiencies are large enough such that they may be practicably engineered into display design.

The separate scientific literatures on STM and cognitive workload are reviewed below to demonstrate the vital link between these two information processing concepts. Two experiments that investigated the feasibility of using stimulus class as an information portrayal variable for reducing STM demands are presented next. And, finally, conclusions bearing on the design of information displays are offered.

SHORT-TERM MEMORY AND WORKING MEMORY

The Capacity Issue

Short-term memory and working memory (WM) are theoretical constructs defined in terms of storage and processing capacities, respectively. Klatzky (1980) analogized them to the space available on a carpenter's workbench. The space may be used for work (i.e., WM) or storage (i.e., STM). Allocation of resources to one memory process reduces the available STM space for the other process. STM and WM often are limiting variables in the human operation of aviation systems.

The limited capacity of STM essentially sets an upper limit on the amount of flight and mission information that crew members can retain during any moment. Over time, this information must be rehearsed or processed by the individual to avoid interference and decay. This issue is critically important because operations that encroach upon memory limitations should be avoided. Additionally, while it is generally accepted that 7±2 unrelated units (or chunks) of information can be stored in STM (Miller, 1956), it is unclear as to exactly the quantity of information that constitutes one chunk. Recall that a chunk of information is based on a rule or correspondence among items that result in a familiar pattern. A chunk, therefore, may be a digit, word, or piece of flight data. Because the size of a chunk itself is indeterminate, it is difficult to convert the capacity concept into a memory loading metric.

To apply STM concepts to the design of information displays, it is constructive to approach STM from the perspective of *efficiency* instead of *capacity*. For example, the exploitation of chunking strategies (e.g., parsing, associating with information held in long-term memory, or capitalizing on familiarity) are considered efficiency issues because the objective is to demand less of STM resources while processing a larger quantity of information per unit time. In the high CWL scenarios of military operations, STM efficiency is a critical issue affecting successful task completion.

The Efficiency Issue

Sternberg (1969) showed that the rate at which STM items are processed or scanned in a choice reaction time task varies as a function of stimulus class. Within a stimulus class, processing time generally increases as a linear function of the number of items to be processed. For example, Sternberg (1966) found that it took 38 ms to process each additional digit held in STM. Since that classic work, other psychologists have manipulated a variety of stimulus classes using the Sternberg paradigm. This paradigm involves the presentation of a series of items known as a memory-set (MSET). Following presentation of the MSET items, a probe or target item is shown to the observer who indicates whether it was one of the MSET items. Items used in an MSET usually are drawn from a single stimulus class, such as visual and auditory digits and letters, familiar shapes, colors, words of various lengths, pictures of common objects, and phonemes (Sternberg, 1975).

Stimulus class processing rate. Klatzky, Joula, and Atkinson (1971) utilized the Sternberg paradigm to measure processing rate differences between letters and animal pictures. A variety of testing conditions were manipulated, but we consider only the condition that included exclusively letters or exclusively picture items. Klatzky et al. reported processing rates of 44.3 ms and 72.7 ms for letters and pictures, respectively. In another study, Burrows and Okada (1971) tested digits that were presented either in a fast condition (MSET items shown for 500 ms each) or a slow condition (MSET items shown for 1200 ms each). Processing rates of 23.7 ms and 30.1 ms were obtained for the slow and fast conditions, respectively.

Cavanagh (1972) compiled data from 32 experiments that investigated processing rate using the Sternberg paradigm. Additionally, data from 13 other studies that measured memory spans were also reviewed. The data from both indices showed a highly correlated linear relationship between processing rate and the reciprocal of memory span (see Figure 1). Cavanagh found that the faster the processing rate for a stimulus class, the larger its memory span. The class of digits, in particular, exhibited the fastest processing rate and also the largest memory span. Digits, therefore, may be classified as the "most efficient" stimulus class in comparison to letters, colors, words, and geometrical shapes.

Inspection of Figure 1 suggest a remarkably simple relationship between stimulus class and processing rate that has potential application in crew station display design. For example, crew members are presented with various classes of stimulus information -- flight and systems status information depicted as words, shapes, letters, and digits. One possible strategy for optimizing the processing of such information is to use the stimulus class which incurs the shortest processing time *if* the information readily lends itself to portrayal using this class and *if* class translation is not required. For example, if the only requirement is to differentiate ground objects from air objects, Figure 1 suggest that colors, as opposed to iconic or verbal representations, should be used. Actually, practical recommendations are not as straightforward since stimulus class optimization is a much more complicated matter.

Note that Cavanagh's (1972) conclusions cannot be viewed as sufficient evidence for the STM differential processing efficiency concept due to several reasons. First, procedures varied widely in the studies reported. For example, Brener's (1940) stimuli were ink-drawn on 7.5 cm x 12.25 cm notecards presented manually for 2 s each, while Joula and Atkinson (1971) utilized tachistoscopic stimuli and had subjects memorize the MSET items by repeating them out loud. This difference could have affected encoding

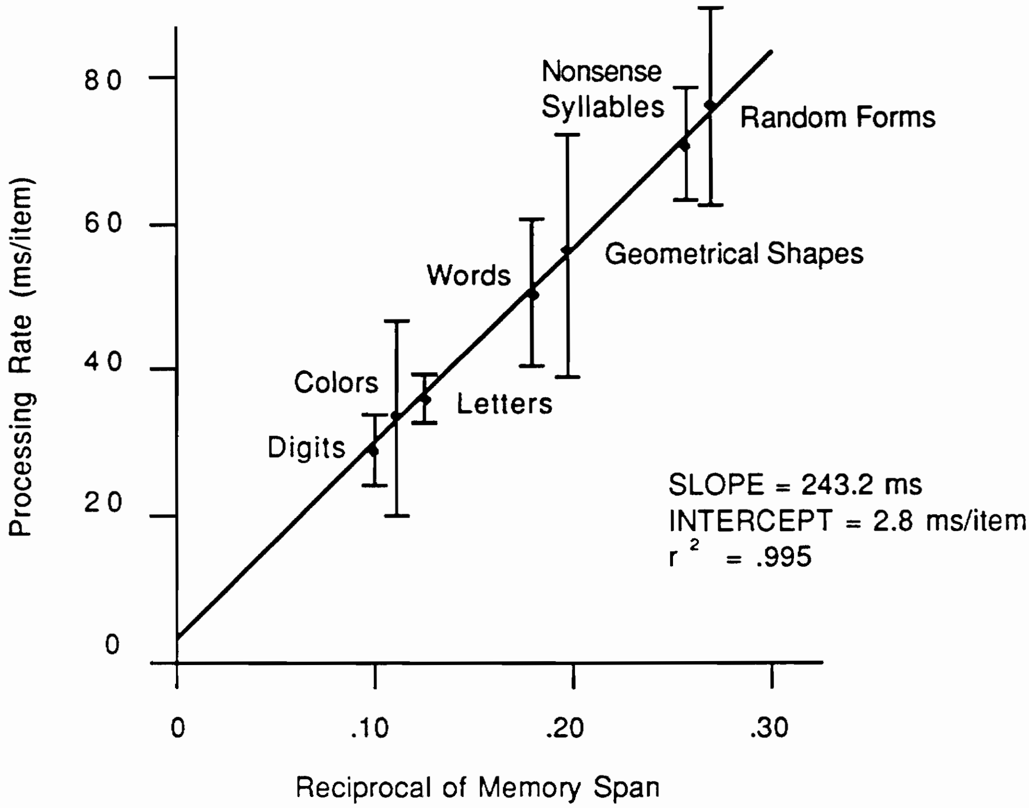


Figure 1. Processing Rate and Memory Span as functions of Stimulus Class (adapted from Sternberg, 1975).

strength and, consequently, recognition latency. Second, Cavanagh did not differentiate between studies that used fixed MSETs (same MSET items across trials) and those that used varied MSETs (different MSET items across trials). Because the former condition has been shown to result in a faster processing rate than the latter condition (Egeth and Smith, 1967; Smith, 1967), Cavanagh's conclusions may have been contaminated. Third, Cavanagh analyzed processing rate data derived from the first two sessions of each study only. This approach potentially is damaging because one is left without knowledge of how practice might have interacted with the experimental effects. As we shall now see, practice effects -- which relate to the notion of automatic versus controlled processing -- in recognition time experiments can substantially alter processing time especially across MSET size.

Automatic and controlled processes affecting processing rate. Schneider and Shiffrin (1977) had subjects process sets of letters or digits using a methodology similar to the Sternberg paradigm. First, one, two, or four MSET items were committed to memory. Then, a series of frames also containing one, two, or four items was shown to subjects. These frames were presented for specific durations ranging from 40 ms to 800 ms. Two experimental conditions were employed. In the consistent mapping (CM) condition, MSET and frame items were drawn from different stimulus classes. In the varied mapping (VM) condition, MSET and frame items were from the same stimulus class. Items that were part of both the memory and frame sets were called targets while frame items not part of the MSET were called distractors. The authors found that in the VM condition, accuracy performance was strongly affected by increases in MSET and frame sizes. In the CM condition, however, accuracy was unaffected by MSET and frame sizes. In other words, response time was not affected by the number of target or distractor items presented when they belonged to different stimulus classes. The term *automatic processing* was used to describe this fast, parallel, and apparently effortless STM process. In contrast, response time increased with the number of target and distractor items when they belonged to the same stimulus class. The term *controlled processing* was used to characterize this slow, serial, and effortful STM process.

In a follow-up study, Shiffrin and Schneider (1977) employed a hybrid of the VM and CM conditions. Only letters were used, but MSET letters were different from frame letters. The authors found that after extensive practice (more than 2000 trials), performance in this condition gradually improved to match performance in the CM condition of the Schneider and Shiffrin (1977) study. Subjects responded equally fast to the frame items whether the MSET size was one or four.

To generalize Shiffrin and Schneider's (1977) findings, the processing rate relationship shown in Figure 1 may be different if different amounts of practice were associated with each stimulus class. Also, if one concurs with the statement that practice has a differential effects on different stimulus classes, then it becomes difficult to accept Figure 1 at face value. In other words, how stable are the processing rate and memory span differences?

Individual differences affecting processing rate. Figure 1 shows that the standard deviation of the mean for all stimulus classes is large enough such that there is a great deal of overlap. These variations may be due to several reasons. First, they be due to the slight methodological differences of the various studies reported. If this is the case, the general between stimulus class relationship would still hold. Second, the variations may reflect actual individual differences in which case the general relationship is nonexistent at least for a heterogeneous population with different levels of STM ability.

To utilize stimulus class as an information optimization technique, the source of the variation must be identified. A means towards this end would be to test different stimulus classes in the same experiment, thereby achieving control over all experimental and extraneous variables. Brown and Kirsner (1980) did just that. These authors replicated Cavanagh's (1972) findings for the stimulus classes of letters, digits, words, and nonsense syllables. Their results are depicted as part of Figure 2.

As Figure 2 shows, both studies reported the same general relationship. Brown and Kirsner (1980) reported that the superior processing rates of their study was probably due to the greater amounts of practice given to their subjects. Again, however, wide variations in processing rates were found. The authors, therefore, decided to analyze the data for individual differences -- subjects were drawn from the undergraduate population of a local university and included 14 females and 17 males aged between 18 and 27 years. The authors reported, "The [slope] functions for the individual subjects, however, bore little resemblance to the group functions, even among those subjects with high correlations between [memory] span and processing rate" (p. 182).

Other factors affecting processing rate. In addition to the practice or familiarity effect, a number of other variables also affect processing rate. For example, Lachman, Lachman, and Butterfield (1979) examined the tradeoff between processing rate and accuracy. They argue that processing rates have to be interpreted with respect to the level of accuracy achieved. In the case of Sternberg's (1975) results, the authors wrote, "People would have to have exactly zero errors at the minimum reaction time in order for Sternberg to infer the precise rate of memory scanning" (p. 163).

Another factor shown to influence processing rate is the serial position of the target item in the MSET. An experiment by Sternberg (1967) demonstrated this effect. Sternberg had subjects memorize from three to seven digits presented serially. He found that the time it took to respond correctly to the probe item varied as a linear function of the probe's

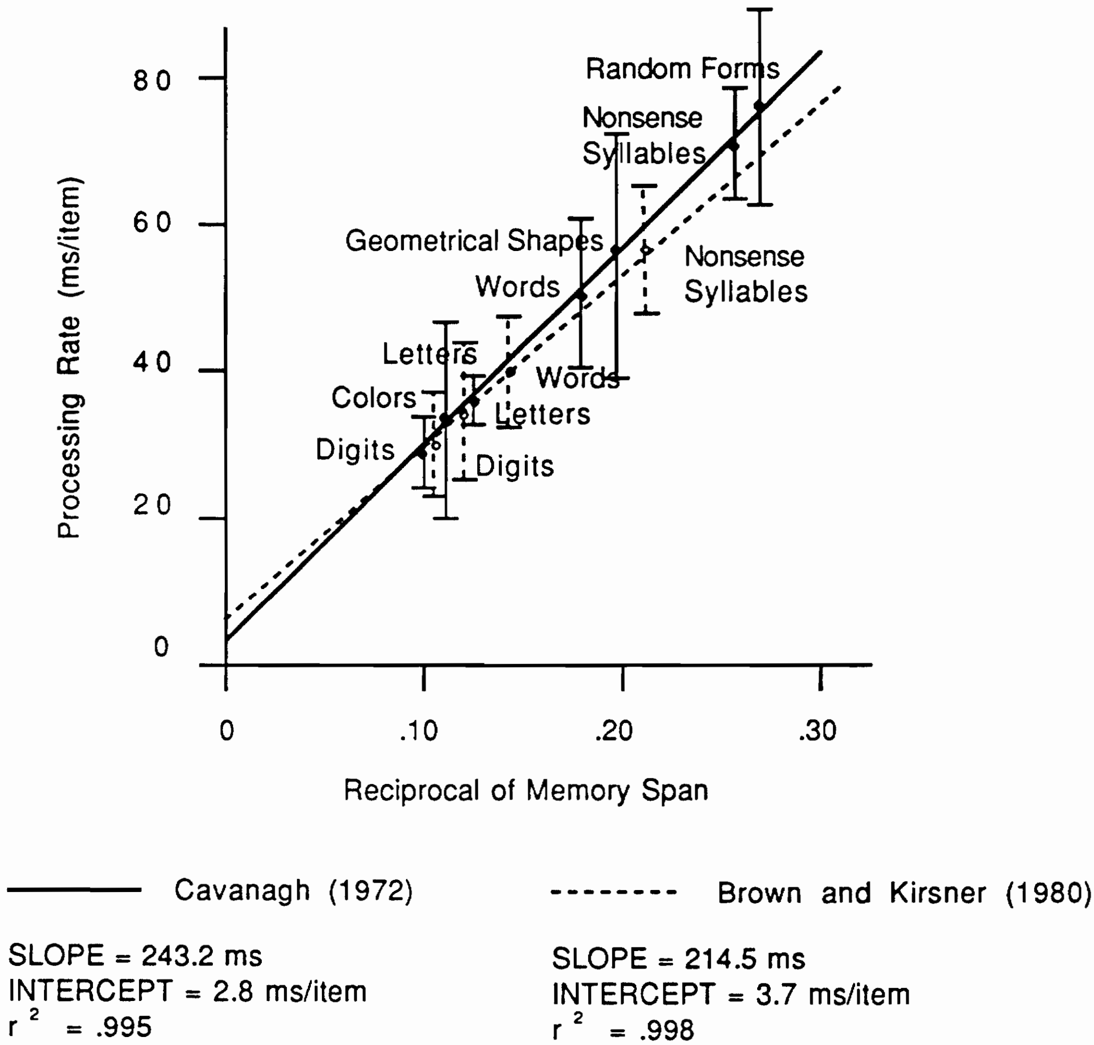


Figure 2. Processing Rate and Memory Span as functions of Stimulus Class (adapted from Brown and Kirsner, 1980, and Sternberg, 1975).

position in the memorized list. A similar serial position effect was also reported by Burrows and Okada (1971).

Thus far, the discussion has been limited to the manipulation of experimental variables (presentation rate, practice effects, etc.) and how these variables alter the stimulus class processing rate relationship shown in Figures 1 and 2. It is probably safe to say that these phenomena are largely confined to the laboratory environment. For example, the processing of stimulus class information in military target search and acquisition operations can hardly be considered purely automatic; neither are they purely controlled. The emphasis on recognition accuracy and target serial position would likely be constant across different stimulus classes used in information displays. No mention has been made of how stimulus class items are processed in STM. Two models, in particular, are highly relevant when considering stimulus class structural differences -- Baddeley's working memory model and Wickens multiple resource theory. These models are discussed next.

Baddeley's Working Memory Model

Using the term *working memory*, Baddeley and Hitch (1974) proposed a model to describe the functional role of STM in information processing. The model consists of a central processing unit or *central executive* serving a number of secondary slave systems. The *visuo-spatial scratch pad* (VSSP) subsystem is thought to be used for the rehearsal and processing of high imagery and spatial information, while the *articulatory loop* is conceptualized as a speech-based subsystem used for rehearsing subvocal or verbal material. A depiction of this model is shown in Figure 3.

This model extends previous STM models to account for the differential effects of stimulus type on memory loading. In particular, it explains the following phenomena (Baddeley, 1984, Chap. 5):

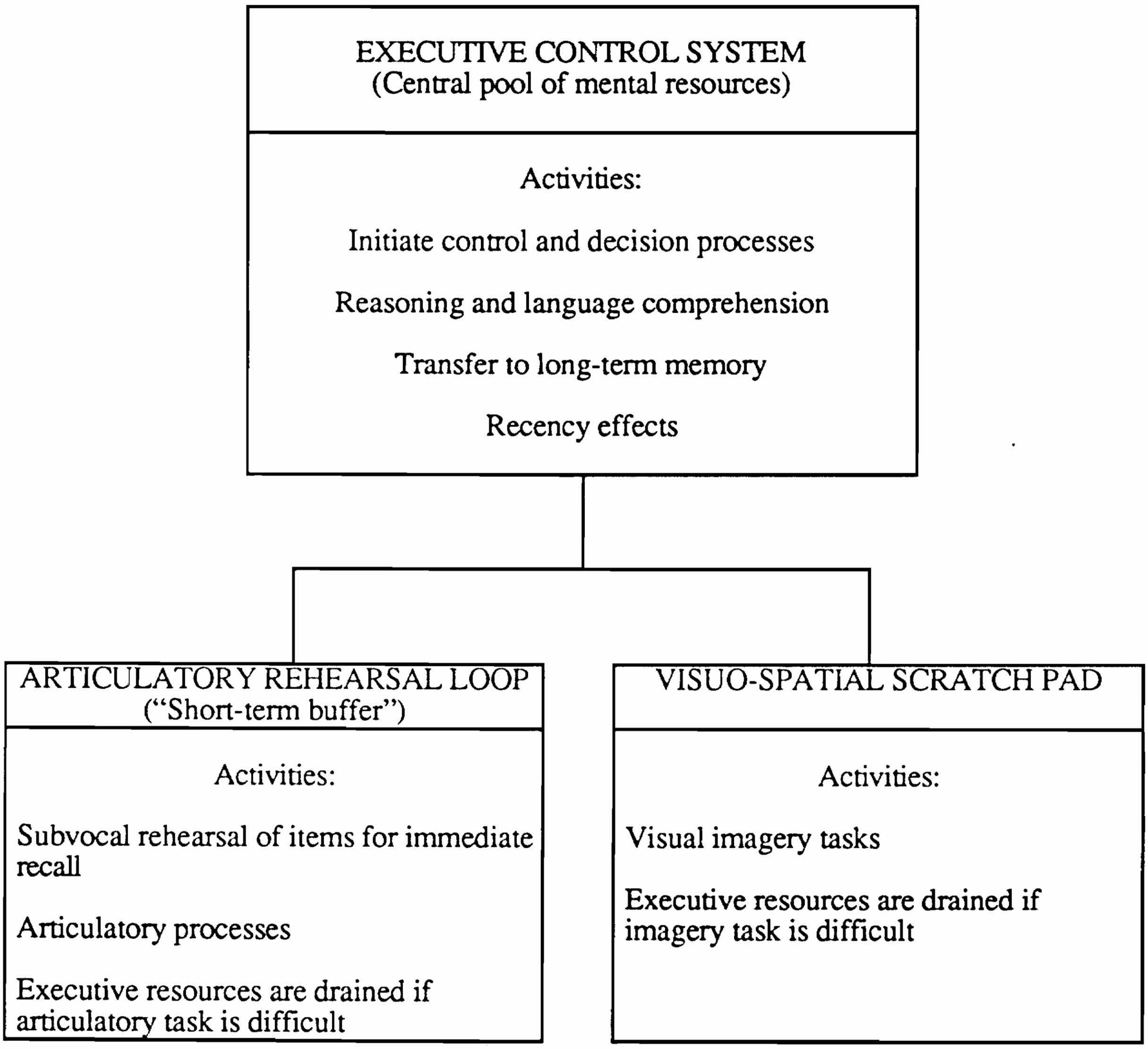


Figure 3. A schematic of Baddeley's Working Memory model.

1. The *phonological similarity effect* where the use of phonologically similar MSET items result in poorer recall than phonologically dissimilar MSET items;
2. The *word-length effect* where immediate recall for word sequences decreases as a function of the time it takes to vocalize the words;
3. The *articulatory suppression effect* where prevention of subvocal rehearsal impairs immediate memory performance; and
4. The *unattended speech effect* where the recall of visual material is impaired by the simultaneous presentation of irrelevant spoken material.

According to Baddeley (1984), the VSSP and the articulatory loop are thought to time-share the processing resources of the central executive. Performance on a mental arithmetic or high imagery task reduces resources available to the VSSP and, thus, interferes with the full functioning of the visual imagery system but not the articulatory loop (Baddeley, 1982). Evidence in support of this statement stems from dual-task experiments in which one task involves STM operations while the other task involves visual tracking. For example, Baddeley, Grant, Wright, and Thomson (1975) had subjects listen to and recall either visuo-spatial word sequences or nonsense word sequences while concurrently performing visual pursuit tracking. The authors found that recall performance was poorer for the visuo-spatial words than for the nonsense words. A similar interference effect also was reported by Baddeley and Lieberman (1980). In an effort to ascertain the locus of this interference, Hitch (1984) provided evidence showing that the VSSP uses a covert mechanism, possibly analogous to eye movements.

Baddeley's WM model is presented because of its relevance to the differential processing efficiency concept in aviation systems operation. In the Gibsonian tradition, colors and shapes afford spatial representations while letters, digits, and words afford verbal representations. Because the laboratory tracking task -- requiring spatial resources

-- is analogous to aircraft handling, the above studies suggest that in order to reduce task interference for manual piloting, the information that pilots receive should be coded in a verbal rather than a spatial format.

Wicken's Multiple Resource Theory

Another model of STM resource allocation is the *multiple resource theory* developed by Wickens (1980). This theory proposes three dichotomous dimensions that may be used to describe differences in mental resource competition. The first dimension differentiates perceptual/cognitive activities from response processes. The second dimension states that analog/spatial activities draw resources different from verbal/linguistic activities. The third dimension defines resource competition based on the mode of processing (auditory vs. visual input, and manual vs. vocal response). The structure of this model is depicted in Figure 4.

Wicken's model is one that is capable of predicting performance when two or more tasks have to be time-shared. The partitioning of resources along the three dimensions attempts to explain how efficiently this may be accomplished. To the extent that both tasks share common levels on any of the dimensions, a difficulty-performance tradeoff will occur; to the extent that they do not, the performance of one task will be insensitive to increased demands from the other task (Wickens, 1980).

A study by Frick (1984) found evidence in support of this model when he attempted to increase the memory span for digits. Frick demonstrated that the memory span increased three digits over baseline when both visual and auditory input modalities were used. Tsang and Wickens (1984) found that two tasks could be better time-shared when response processes were logically allocated between motor and speech outputs than when only motor

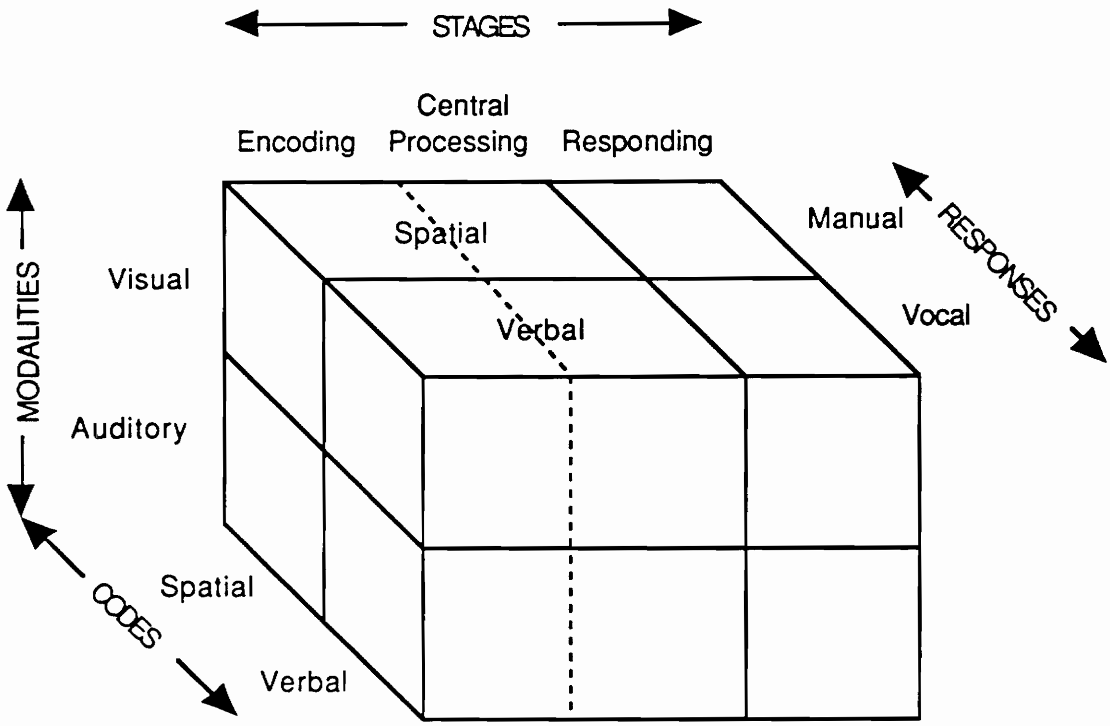


Figure 4. The structure of Multiple Resources (adapted from Wickens, 1984a).

outputs were required. Two tasks using different memory systems will be carried on in parallel more efficiently than two tasks using a common memory system (Wickens, 1989).

Baddeley's WM model and Wickens multiple resource theory have important implications for the design of information displays, especially in elevated-CWL situations. Both predict time-sharing efficiency for two or more STM tasks performed concurrently. As resources along more dimensions are time-shared, there is a resulting decrease in time-sharing efficiency and an increase in difficulty-performance tradeoffs (Wickens, 1989).

It is important to note that neither model is fully developed to account for all types of STM task interference. Wickens (1989) cited evidence of other resource systems such as the *visual subsystem* with resources devoted either to foveal (object recognition) or peripheral (spatial orientation) images. Baddeley (1984) wrote, "...although I believe we know rather less about visual than phonological coding in memory, we already know a great deal, and have available techniques which I believe will make it possible within the next few years to obtain a much more coherent view of the relationship between the clusters of evidence at present available" (p. 144).

COGNITIVE WORKLOAD

As in STM work, CWL research has received sustained attention due to the shift from mainly physical work to work that demand more of our higher cognitive processes. For the aviation community, in particular, the Federal Aviation Administration is considering aircraft certification in terms of workload metrics, and the U. S. Air Force may impose workload criteria on future aircraft designs. Workload metrics could be used for such purposes as (1) comparing design alternatives in terms of imposed workload, (2) determining workload criteria for flight routines, (3) selecting crew members according to their workload capacities, and (4) allocating tasks between crew members and the aircraft (Sanders and McCormick, 1987).

In a review of the major workload conferences and reports, O'Donnell and Eggemeier (1986) wrote that most CWL scientists follow fairly well-defined paths in formulating and assessing operational definitions and metrics for measuring cognitive effort expenditure in specific situations. It must be mentioned, however, that despite two decades of research, a precise and acceptable model of the CWL theoretical construct still does not exist (Reid and Nygren, 1988; but see Meshkati, 1988, for developments towards a cohesive model). Furthermore, a universal metric of CWL applicable to a wide variety of situations has been even more elusive.

Given the multidimensional nature of CWL, it is not surprising that scientists from many disciplines have contributed to this field. They include physiologists who investigate workload in terms of central nervous system changes that take place involuntarily in response to changes in CWL level (Wierwille, 1979). There are systems engineers who take a mathematical modelling approach by assessing system (machine and human) inputs and outputs (Rouse, 1979). And, control engineers trace the dynamics of a system and have contributed with a fairly well-standardized set of manual tracking tasks, useful especially for perceptual-motor workload assessment (Jex and Clement, 1979). From these and other approaches, four major categories of CWL assessment techniques have evolved (1) physiological-based measures, (2) subjective-based measures, (3) primary task-based measures, and (4) secondary task-based measures. Each will be briefly discussed as it relates to the methodologies used in the present experiments.

Cognitive Workload Assessment Categories

Physiological-based measures. The basic quest of physiological measures may be summarized in the question: What are the physiological indices that occur in response to changes in CWL level? For example, pupillary dilation has been found to vary as a function of the number of digits held in STM (Beatty, 1982), while sinus arrythmia (specifically, the 0.10 Hz component) has been shown to decrease systematically as more letters were held in STM (Aasman, Mulder, and Mulder, 1987). Thirteen of the more common physiological indices used in CWL research include galvanic skin response, muscle tension, pupillary dilation, electrooculogram, body fluid analysis, speech pattern analysis, respiration analysis, evoked cortical potential, flicker fusion frequency, heart rate, electrocardiogram, electromyelogram, and electroencephalogram (Wierwille, 1979). In addition, there have been attempts at using relatively unconventional approaches such as psychoneuroendocrine analysis, alveolar gas concentration level, and deep auditory canal temperature (see citations by Meshkati and Loewenthal, 1988).

The history of physiological indices has convinced most researchers that a single technique is unable to provide an adequate overall workload assessment (O'Donnell and Eggemeier, 1986). Many physiological techniques suffer from the disadvantage of necessitating physical intrusion (i.e., equipment attachment) during task performance.

Despite this limitation, Wilson and O'Donnell (1988) have successfully developed a Neuropsychological Workload Test Battery (NWTB). According to the authors, the NWTB was derived from an aggregate of several extant techniques and it can differentiate among several information processing mechanisms. The authors, however, do recommend using additional subjective and/or performance measures to provide exhaustive answers to different CWL questions.

Subjective-based measures. Subjective measures seek to infer workload level by obtaining the opinions, judgments, and estimations of the person performing the task. They have been used extensively due to several practical advantages such as ease of administration and low cost of implementation. Subjective measures are relatively non-intrusive and they afford a high degree of face validity. O'Donnell and Eggemeier (1986) noted that the theoretical basis for their use is grounded on the assumption that "increased capacity expenditure...will be associated with subjective feelings of effort or exertion that can be accurately reported by the subject" (p. 42.7).

Hart and Wickens (1990) suggested that the following questions should be answered in selecting the subjective technique(s) to be used (1) the focus of the research question, (2) the grain of analysis required, (3) the probable level and sources or workload, (4) the importance of diagnostic information, and (5) the practical constraints imposed by the research environment. The authors categorized subjective assessment techniques as uni-dimensional (e.g., free modulus magnitude estimation, bisection), hierarchical (e.g., Copper Harper, Bedford), or multi-dimensional (e.g., NASA-Task Load Index [NASA-TLX], Subjective Workload Assessment Technique [SWAT]). Within the context of STM/CWL research, multi-dimensional techniques have been widely tested and found to be comparatively more sensitive; uni-dimensional and hierarchical techniques, useful in other applications, do not provide the degree of diagnosticity available from multi-dimensional techniques (Hart and Wickens). In the following paragraphs, in-depth discussions of SWAT and the NASA-TLX, including their strengths and limitations, are presented.

The development of SWAT was based on the following requirements (Reid and Nygren, 1988, p. 188):

1. Development of a precise model while minimizing the physical intrusiveness of the data collection process on the operational situation;
2. To ensure minimal measurement constraints on the complexity of the subjective workload judgment required of the operator; and
3. To provide a means of testing the validity of SWAT and its underlying psychometric principles.

In order to capture the multidimensionality of CWL, SWAT's developers sifted the literature to arrive at more than 20 workload definitions. Based on concordance and sometimes disagreement among these definitions, three dimensions which operationally define CWL were identified. These dimensions are (Reid, Potter, and Bressler, 1989, Appendix B):

1. *Time Load.* This is the feeling of time pressure for task completion. It includes the perception of both free time (time available exceeds time required to perform task) and task overlap (when two or more tasks compete for the same time resource and tradeoffs must be made).
2. *Mental Effort Load.* This accounts for task factors such as difficulty, complexity, and the perceived effort involved in problem solving and decision making. Sensory information demands are also included.
3. *Psychological Stress Load.* This encompasses operator variables such as emotional state, motivation, health, and fatigue. It reflects anything that contributes to

operator anxiety, frustration, and/or confusion.

Three levels of each dimension are used in the SWAT scale (see Table 1). The authors of SWAT emphasized that their objective was not to end all theoretical debates on the definition of CWL. Rather, they "intended to capture most of the important components that appear to influence people's perception of workload. SWAT is intended to be a pragmatic approach to the estimation of cognitive workload in operational situations" (Reid and Nygren, 1988, p. 191).

Two distinct steps are involved in the use of SWAT. The first step is scale development in which 27 workload levels (three dimensions with three levels each; 3^3) are rank ordered by individual operators. The ranking is subjected to conjoint measurement and scaling procedures to arrive at a transformation scale with interval properties. This scale runs from 0 to 100 with 0 and 100 mapping to the 1,1,1 and 3,3,3 combinations, respectively. An illustration of the three-to-one dimension mapping is shown in Figure 5.

The second step in using SWAT is event scoring. For a given task, the operator rates the workload experienced by assigning a level from each of the three dimensions. Laboratory experiments using SWAT typically require subjects to rate the workload associated with performing a series of trials. The ratings, after conversion to the 100-point interval scale, are analyzed via parametric statistical techniques.

The major benefit of using SWAT is that while other subjective techniques might require interval- or even ratio-scaled judgments (e.g., free modulus magnitude estimation over several workload dimensions or the use of bisection estimation), SWAT only requires ordinal judgments. It is the conjoint scaling of operators' judgments that results in the interval scale. It is, therefore, a more powerful technique amendable to parametric analysis -- more powerful than the NASA-TLX technique which gives only ordinal data. Used in

TABLE 1

The Three SWAT Dimensions (after Reid, Potter, and Bressler, 1989)

I. *Time Load*

1. Often have spare time. Interruptions or overlap among activities occur infrequently or not at all.

2. Occasionally have spare time. Interruptions or overlap among activities occur frequently.

3. Almost never have spare time. Interruptions or overlap among activities are very frequent, or occur all the time.

II. *Mental Effort Load*

1. Very little conscious mental effort or concentration required. Activity is almost automatic, requiring little or no attention.

2. Moderate conscious mental effort or concentration required. Complexity of activity is moderately high due to uncertainty, unpredictability, or unfamiliarity. Considerable attention required.

3. Extensive mental effort and concentration are necessary. Very complex activity requiring total attention.

III. *Psychological Stress Load*

1. Little confusion, risk, frustration, or anxiety exists and can be easily accommodated.

2. Moderate stress due to confusion, frustration, or anxiety noticeably adds to workload. Significant compensation is required to maintain adequate performance.

3. High to very intense stress due to confusion, frustration, or anxiety. High to extreme determination and self-control required.

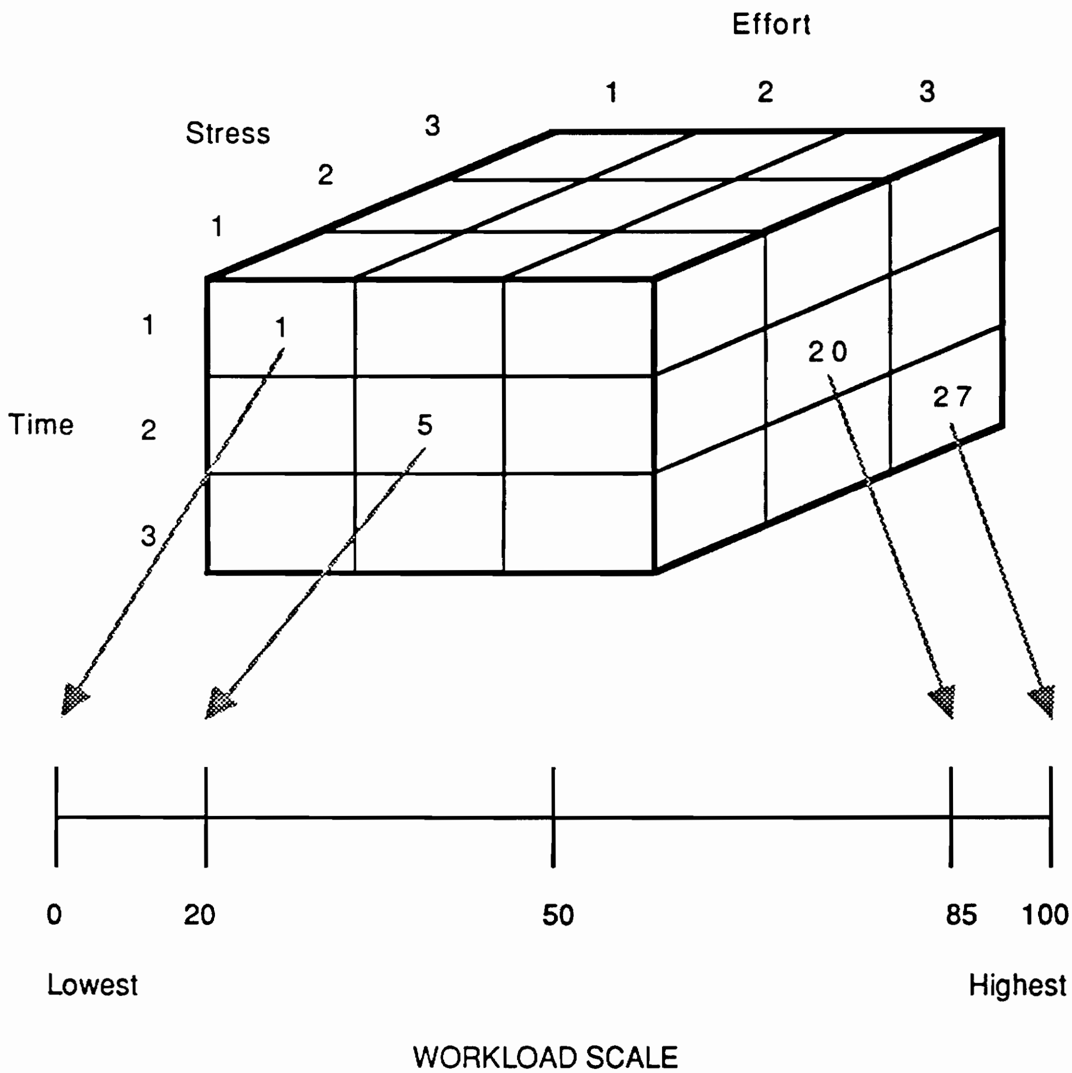


Figure 5. SWAT and the use of Conjoint Analysis to Rescale individual ranks to a unique Interval Scale (adapted from Lysaght et al., 1989).

STM investigations, it is certainly desirable for measuring the amount of spare mental capacity or "free slots" according to Miller's 7±2 magic number.

Conjoint analysis performs a series of axiom tests on the 27 workload ranks. These tests determine how well the rank orderings conform to one of the following four polynomial models: additive, multiplicative, distributive, and dual-distributive. A high number of test failures indicates that the ranks do not conform to an interval scale transformation. After conjoint analysis, the ranks are rescaled through use of the MONANOVA (Kruskal, 1965) and NONMETRG (Johnson, 1973) scaling algorithms. These algorithms determine a best-fitting line based on the least-squares approach commonly used in regression analysis. Measures of THETA and STRESS (between 0.0 to 1.0) are provided to indicate how well the ranks fit.

Depending on the SWAT analysis objectives, the ranks may be scaled for each operator or for groups of individuals. When used for group analyses, Kendall's Coefficient of Concordance (W) can assess the amount of inter-individual agreement. The group scale has the advantage of averaging individual idiosyncrasies while single operator scales reveal individual differences in workload perception. The Time, Effort, and Stress dimensions may be separately examined. For example, a high rating for the time dimension only might indicate an inefficient task sequencing or bottlenecks in flight operation.

SWAT has been tested extensively in aviation operations. Rank orderings produced as far apart as one year have been found to correlate better than 0.90 (Reid and Nygren, 1988). Relatively high Coefficients of Concordance have been obtained from rank orderings involving homogeneous groups of individuals (Kendall's W of between 0.76 and 0.82 reported by Reid, Shingledecker, and Eggemeier, 1981).

Since its initial development, there have been several refinements and modifications to SWAT. For example, Pro-SWAT (Projective SWAT) has been modified for use in the preliminary phases of systems design (Eggleston and Quinn, 1984), while SWAT 2 and SWAT 3 have increased measurement sensitivity for low workload situations (Reid, 1985).

Similar to SWAT, the NASA-TLX scale estimates overall workload based on the aggregate ratings of several dimensions (Hart and Staveland, 1988). Six subscales (Mental Demand, Physical Demand, Temporal Demand, Own Performance, Effort, and Frustration) are rated from 1 to 100 points in 5-point increments. In addition, the relative importance of each subscale is rated on a scale from 0 to 5. The resulting workload score is the weighted average of the subscale ratings multiplied by their respective weights. Figure 6 graphically demonstrates the rating procedure for this workload assessment technique.

The NASA-TLX scale has been used successfully in laboratory and applied environments (e.g., aircraft simulators, helicopter handling, ground-based systems). Hart and Wickens (1990) noted that it provides consistently lower between-rater variability compared to other subjective scales. It has also better diagnostic capabilities when compared to SWAT.

Primary task-based measures. Primary task measures, which seem to be the most direct means of workload assessment, specify performance adequacy on the principal task of interest. They are based on the assumption that the quality of operator performance changes as a direct result of changes in workload level (Williges and Wierwille, 1979). Primary measures are popular because there is little need for measurement equipment setup or physical intrusion during task performance. Data can be recorded automatically and analyzed at a later time.

SUBSCALE	TASK RATING	RATING		WEIGHT		PRODUCT
Mental Demand	|_x_____________| Low High	10	x	3	=	30
Physical Demand	|_____x_________| Low High	35	x	0	=	0
Time Demand	|_________x_____| Low High	55	x	5	=	275
Own Performance	|____x__________| Low High	30	x	1	=	30
Effort	|_____________x_| Low High	80	x	3	=	240
Frustration	|___x___________| Low High	20	x	3	=	60
		Sum			=	635
		Total Weights			=	15
		Mean Workload Score			=	42

Figure 6. Workload Rating Procedure for the NASA-TLX Scale (adapted from Hart and Staveland, 1988).

Primary tasks may be performed singly or in combinations of two or more tasks. When used in a dual-task paradigm, a secondary task serves either as a loading or subsidiary task. The use of dual-task paradigms depends on the nature of the research objectives. For example, the paradigm used in some STM studies investigating spare cognitive capacity requires individuals to maintain adequate performance on the primary task. Only if spare resources were available they would be devoted to the secondary task.

The major types of primary measures include reaction time, error rate, frequency of task performance, and total time on target if a tracking task was used. To be effective, Meister (1985) recommends that primary measure(s) should be chosen to reflect maximally manipulation of the variable(s) of interest.

A major deficiency of primary measures is that they lack sensitivity in detecting changes in CWL level (Meshkati and Loewenthal, 1988). Williges and Wierwille (1979) concluded that low operator workload conditions may not be perceptible by primary measures since operators can adapt in an effort to maintain a constant output; high workload situations, however, are better detected. Another deficiency of primary task measures in applied settings is their high task specificity and, consequently, their interpretation difficulties for between-task comparisons.

Secondary task-based measures. In order to optimize system performance, the limitations and capacities of the human operator need to be determined. Secondary measures of CWL may be used for this purpose. The logic behind the use of secondary task measures is that spare capacity available during performance of a primary task can be consumed by introducing a secondary task. The aim is to identify the point or region of maximal performance beyond which greater demand by one task will result in decreased performance of the other task. To illustrate this logic, suppose the primary task is aircraft handling under different loading conditions (e.g., low level flying under clear or foggy sky). Also suppose that the secondary task is digit recognition using Sternberg's paradigm. Under the clear sky condition, one would expect a larger digit span because more STM resources would be available to the pilot.

As the example above demonstrates, secondary measures assume that overall processing or channel capacity is fixed across all test levels and testing conditions. It is also assumed that both primary and secondary task efforts are obtained from a common resource pool (Wickens, 1984a). Thus, workload levels for both tasks should be linearly additive. With linear additivity, measures of CWL level can be specified on an ordinal, if not an interval scale. Secondary task paradigms, by definition, always involve two or more tasks.

The preceding discussion of memory capacity and efficiency in workload assessment research approached STM and CWL as separate issues. The next section discusses recent research that show these two issues to be vitally linked.

SHORT-TERM MEMORY/ COGNITIVE WORKLOAD RELATIONSHIP

Short-term memory, whether used for storage, processing, or transfer of information to and from long-term or sensory store, is a crucial element for flight success in the highly automated cockpits of modern aircraft. As a consequence, attempts have been made to measure STM loading in a wide variety of single- and dual-task conditions. According to Reinhart et al. (1988), STM tasks in CWL research accomplish two purposes. First, they provide a source of CWL that can be controlled easily through the manipulation of stimulus presentation rate and duration, the number of stimuli to be presented, decay interval, etc. Second, STM tasks provide their own performance measures of CWL (e.g., response latency, error rate).

Studies Supporting the Relationship

Eggemeier, Crabtree, Zingg, Reid, and Shingledecker (1982) conducted a Sternberg-type sequential STM task to evaluate the sensitivity of SWAT to various stimulus presentation rates and different MSET sizes. The authors concluded that their study provided "very strong support for the applicability of SWAT as a workload measurement procedure in the type of STM task employed" (p. 647). Specifically, subjective impressions of workload correlated significantly with the number of letters held in STM. Reinhart et al. (1988) use a tracking-primary/Sternberg-secondary task paradigm to investigate the effects of digit loading on CWL. Their major finding was that increased memory demands were a major factor in influencing subjective workload ratings.

Other studies that have utilized and provided favorable comments for the SWAT technique in STM/CWL research include:

1. Effects of delayed SWAT reporting in a Sternberg-type STM update task

(Eggemeier, Crabtree, and LaPointe, 1983);

2. Spatial STM (Eggemeier and Stadler, 1984);
3. Spatial STM transformation (Vidulich and Tsang, 1985); and
4. STM perceptual-motor tracking (Gopher, Chillag, and Arzi, 1985; Hemmingway, 1984).

Use of the Sternberg Paradigm in Short-Term Memory/Cognitive Workload Research

Wickens, Hyman, Dellinger, Taylor, and Meador (1986) suggested two reasons for use of the Sternberg paradigm in aviation CWL research. First, as a secondary task, it has high face validity due to its similarity with the communication monitoring task of aircraft crew members which, by definition, is also a task secondary to control of the aircraft. Second, numerous studies have shown that it adequately describes cognitive performance based on a human information processing model that is well-validated (i.e., STM).

CWL scientists have identified properties that determine the validity and reliability of CWL assessment techniques. The discussion now shifts to how the Sternberg secondary task technique fairs in terms of these properties.

Sensitivity. Sensitivity refers to the ability to reveal mental loading at different task demand levels. A metric that can differentiate large workload differences may be useful only for identifying bottlenecks in flight performance; it would likely be of value for precise workload measurement. Several studies reported earlier found high sensitivity ratings for Sternberg's paradigm as a secondary task (e.g., Eggemeier et al., 1982, 1983; Reinhart et al., 1988; Wickens et al., 1986).

Diagnosticity. This is the ability to determine the exact causes of CWL level changes. Development of this property relates to research in human attention and Wicken's multiple resource theory discussed earlier. Wickens et al. (1986) reported three experiments that revealed the diagnostic capabilities of Sternberg's paradigm when used as a secondary task. Their experimental design involved aircraft handling as the primary task and auditory letter recognition as the secondary task. The authors reported the secondary task to be diagnostic of the two phases of flying employed: holding phase and landing phase.

Intrusiveness. Eggemeier (1988) defined intrusiveness as "the tendency to cause unintended degradations in ongoing primary task performance" (p. 49). He cited evidence indicating that intrusion effects arising from the use of subjective measurement techniques (e.g., SWAT) have been minimal. In a sense, this is expected since the standard practice is to administer the ratings after all tasks have been completed or at discrete trial breaks.

Data on intrusion effects arising from the use of dual-task paradigms, however, have been less than encouraging (Eggemeier, 1988). An absence of intrusion in the subsidiary task paradigm (i.e., secondary task load varies while primary task load is held constant) implies that performance on the primary task should be comparable to that when performed singly. Eggemeier reported that systematic evidence regarding its effects are not extensive, but appear related to the type and level of primary task loading. The Reinhart et al. (1988) study reported earlier employed three levels of a primary compensatory tracking task and a secondary Sternberg task. Their analyses revealed increased RMS tracking error with increased memory-set size and increased tracking difficulty.

Despite problems of intrusiveness, the tracking-primary/Sternberg-secondary task paradigm has nevertheless given fairly robust CWL results. Henry Jex and his colleagues (Jex, 1988; Jex and Clement, 1979; Jex, McDonnell, and Phatak, 1966) attempted to optimize the experimental procedures and variables involved in using tracking as a primary task. The authors developed a Standard Subcritical Tracking Task (SCTT) recommended especially for use with the Sternberg paradigm as a secondary task. Specific control parameters of the SCTT were delineated (Jex, 1988). These include recommendations to use a first-order compensatory tracking task and specification of the required sensitivity and range of the control dynamics.

THE PRESENT RESEARCH

The preceding discussion reviewed the separate short-term memory and cognitive workload literatures and, then, addressed the existence of a dynamic relationship between these two human information processing concepts. In relation to the objective of optimizing stimulus class as an information display variable, both concepts deserve central examination. There is, on the one hand, the issue of STM processing rate for items of different stimulus classes; on the other hand, it is also important to consider the CWL associated with these processes. Together, they embody the notion of *differential processing efficiency.*

The papers by Cavanagh (1972) and Brown and Kirsner (1980) provide excellent discussions of stimulus class processing rate. Through these and other findings, CWL researchers have a platform on which to investigate spare mental capacity using, in particular, the Sternberg paradigm. The studies by Reinhart et al. (1988) and Wickens et al. (1986) have been attempts toward this end. In the Reinhart et al. study, digit loading using a Sternberg secondary task served as a good CWL measure. Wickens et al. obtained similar results using an auditory Sternberg secondary task. These measures are quantitative, objective, and performance-based.

Considering the case for qualitative or subjective measures, the NASA-TLX and SWAT scales have been shown to be sensitive and diagnostic instruments that incur very little intrusion. Both techniques have been successfully applied in STM/CWL research. While the NASA-TLX scale is somewhat more accurate than SWAT (i.e., lower between-rater variability), it is decided that SWAT will be employed in the present research. The rationale for this choice has to do with the methodology employed. The SWAT rating procedure incurs less time and effort than the NASA-TLX rating procedure (compare rating a 1, 2, or 3 on three dimensions as opposed to rating 0 to 100 and ranking each of six subscales). Since the participants in this research will be rating the CWL associated with each Sternberg trial and there will be many such trials, use of the more effortful NASA-TLX rating procedure might be a source of CWL in itself. This would potentially create problems in terms of data analysis and interpretation.

Use of SWAT in the tracking-primary/Sternberg-secondary task paradigm has been documented in several studies (e.g., Reinhart et al., Wickens et al.). Compensatory tracking, in particular, has been shown to be a valid and reliable method for loading STM (Jex, 1988).

The present research seeks to identify opportunities and methods to reduce cognitive workload by reducing STM demand. Specifically, the feasibility of using stimulus class as an information portrayal variable was investigated. Two experiments quantitatively and qualitatively compared memory loading as a function of stimulus class. A tracking-primary/Sternberg-secondary experimental paradigm was employed in the first experiment. The second experiment utilized a single task -- a modified version of the Sternberg paradigm -- to load STM. Within the aviation context mentioned in the introduction, the reduction of STM demand and operator workload is highly desirable.

EXPERIMENT 1

Objectives

This experiment was conducted to test, both qualitatively and quantitatively, the differential processing efficiency concept. The objectives were (1) to compare STM loading (in terms of processing rate) as a function of stimulus class, (2) to investigate the feasibility of using SWAT as a subjective method of assessing CWL, and (3) to describe the relationship between the subjective (SWAT) and objective (recognition latency and error) measures.

Experimental Tasks

A subsidiary dual-task paradigm was employed. The primary task was one-axis first-order compensatory tracking administered at a single level of tracking difficulty. Specific guidelines for using this task can be found in Poulton (1974; see use by Reinhart et al., 1988) and Jex (1988). RMS error or the amount of tracking deviation from target over time was used as the measure of overall tracking performance. For the secondary task, Sternberg's two-choice reaction-time paradigm was employed. Wickens et al. (1986) recommended the following set of implementation procedures for its use:

1. Minimize peripheral input-output delays (e.g., present a fixation cross prior to trial, keep fingers on response keys);
2. Avoid using MSET sizes of one or greater than four or five;
3. Avoid using same MSET for more than 20-30 consecutive trials;
4. Vary MSET sizes regularly to avoid practice and fatigue effects;
5. Collect data from multiple subjects; and
6. Avoid extreme loading conditions.

For the current experiment, recognition latency in response to the various MSET sizes was used to measure spare cognitive capacity as participants concurrently performed the tracking task. SWAT was used as a second measure of spare capacity given the sensitivity and diagnostic limits of using only one workload measure.

Methodology

Research participants and the selection process. Ten male students, aged between 20 and 27 years, from the Virginia Tech community were recruited for this experiment. Each was paid $25 for his participation. All participants signed an informed consent contract (see Appendix A) and underwent a two-part screening procedure. The first part tested for contrast sensitivity using the Vistech Vision Contrast Test System. The criterion for this test was a normal or corrected Snellen acuity of 20/25 or better. The second part tested for color deficiency using the Dvorine pseudo-isochromatic plates (Dvorine, 1953). This test consists of two sections. The first section tested participants' knowledge of color names. The second section tested for color deficiency where participants had to view a series of plates. The criterion for this test was a maximum of two incorrect responses. All participants passed the screening tests.

Apparatus and stimulus. All experimental data were collected using an IBM PS/2 Model 80 microcomputer. This machine has an 80386 20-Mhz processor and an 80387 20-Mhz coprocessor. The display used for presenting the experimental stimuli was an IBM PS/2 8514 direct-drive analog color display. It had a maximum addressability of 1024 x 768 pixels (.28 mm dot pitch) covering an active diagonal of 35 cm.

The computer screen was tilted backwards 15 deg from the vertical plane (i.e., viewing angle was 15 deg below the horizontal plane passing through the eyes) and its center was positioned approximately 45 cm from the participants' eyes. Screen center to floor level was 100 cm while screen center to table level was 28 cm.

A two-axis isometric joystick (Measurement Systems, Model 462) was used as the input device for the tracking task while a custom-made two-key response pad was used for collecting secondary task responses. All participants were right-handed. They, therefore, tracked with their right hand and responded to the Sternberg task with their left hand. The index and middle fingers were used to give the "YES" and "NO" responses, respectively.

The SWAT card sort was completed via paper and pencil. Appendix B presents the instructions used for this procedure. Development of the SWAT scale was accomplished using software supplied by Reid et al. (1989).

The experimental tasks and stimuli were generated using C language (Microsoft, Version 5.10) and the HALO '88 Graphics Kernel System (Media Cybernetics, Version 3.0). Figure 7 is a snapshot of a typical experimental trial. The tracking task components of an inverted triangle (representing the participants' input) and a horizontal line (representing the tracking deviation) were situated at the center of the display. All secondary task stimuli were presented at a constant 8 cm above the center of the display.

The following five stimulus classes were tested: digits, letters, geometrical shapes, colors, and words. Figure 8 is a list of the items used. All classes were restricted to eight items each in order to achieve an equal level of familiarity. The selection of shapes and colors was based on familiarity. The word items were chosen from the Battig and Montague (1969) norms for verbal items. The selection dimension used was *familiarity*. Only monosyllable four-letter words were used.

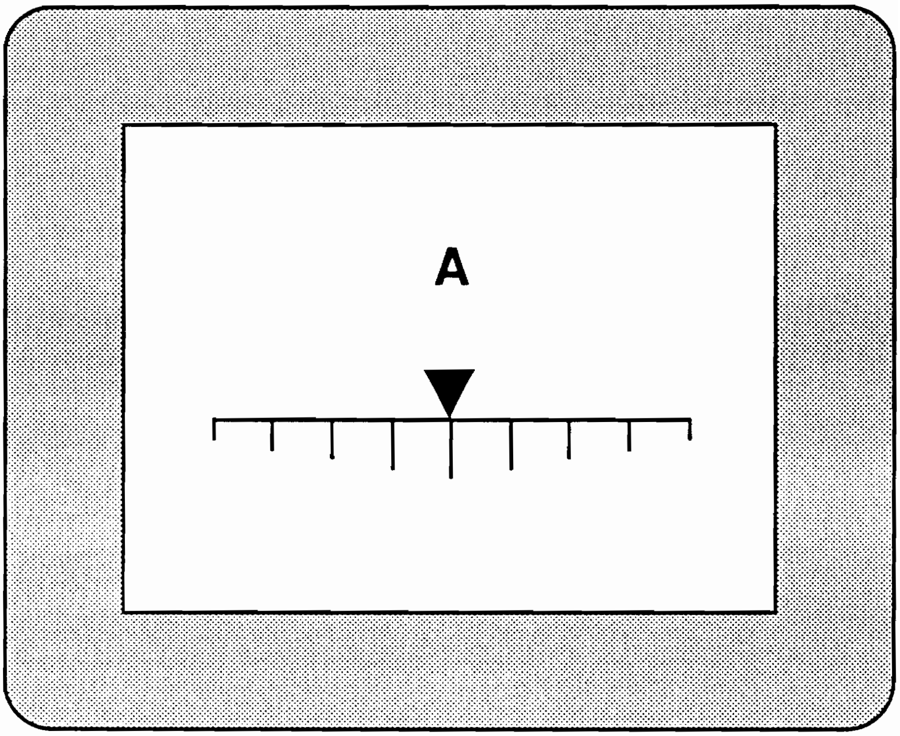


Figure 7. Experiment 1: Snapshot of a Typical Trial.

Letters	A	B	C	D	E	F	G	H
Digits	1	2	3	4	5	6	7	8

Colors	"Red"	"Blue"	"Green"	"Pink"
	"White"	"Gray"	"Purple"	"Yellow"
Words	HAND	GIRL	KITE	SHOE
	DOOR	ROAD	BALL	FROG

SHAPES

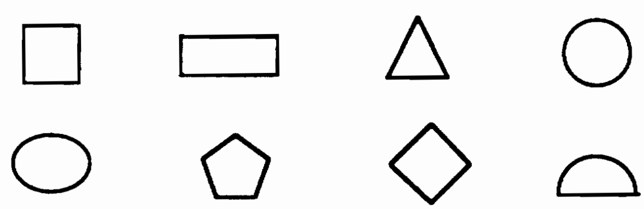


Figure 8. Experiment 1: Stimulus Class Items.

Design and Procedure

The experiment was conducted over five sessions. The first session lasted approximately 2 hr, in which the first hour was devoted to development of the SWAT scale and the second hour was used for administering task instructions and practice on the experimental tasks. Actual data collection occurred on Sessions two through five which lasted approximately 1 hr each.

The tracking task was created by the construction of a random signal. It was administered concurrently with the secondary task. The following RMS-error equation was used to measure tracking performance:

$$\text{RMS error} = [\Sigma(X-X_t)^2/N]^{1/2} \qquad (1)$$

where: X is the cursor or operator tracking position;

X_t is the target position; and

N is the number of samples taken per trial.

The Sternberg secondary task used MSET sizes of two, three, four, and five nonrepeating items. MSET size and stimulus class varied randomly from trial to trial. The probe probability (i.e., percent of trials where participants responded "YES") was set at 50%. Positive and negative probes, therefore, appeared equally often. A typical trial sequence occurred as follows. The tracking task was initiated and 10 s afterwards a fixation crosshair appeared for 1 s signalling that the MSET items were about to appear. The MSET items were presented at a rate of 1.5 s each (80% duty-cycle). After offset of the last item, there was a 2 s interval and the probe item was presented. Participants then gave their "YES/NO" response which terminated the tracking task. The time (in ms) from probe item onset to response constituted recognition latency.

After giving their responses, participants rated the workload associated with the trial. No time limit was set for participants to give their SWAT ratings.

Following the SWAT ratings, a “READY” signal appeared on the screen and participants pressed a “YES” key to initiate the next trial. The instructions given to participants are shown in Appendix C.

Each participant received 400 trials (4 MSET sizes x 5 stimulus classes x 2 probe types x 10 replications) divided into blocks of 40 trials each. The first two blocks served as practice. Actual data were collected on the next eight blocks. Two blocks were run per experimental session and each session was conducted on a different day. There was, therefore, a total of five experimental sessions. Mandatory 5-min rest breaks were imposed between blocks.

Results

Initial analyses focused on the block (or learning) effect. Three analysis of variance (ANOVA) procedures were performed on the independent variables of Block, Stimulus Class, MSET Size, and Probe Type. Tables D1 to D3 present results for the dependent variables of recognition latency (RL), recognition error (RE), and SWAT Ratings, respectively. The Block effect is statistically significant for the dependent variables of RL [$F(7,63) = 7.43$, $p < .0001$] and SWAT Ratings [$F(7,63) = 3.54$, $p = .0028$], but not for RE [$F(7,63) = 0.82$, $p = .5788$]. Post hoc Newman-Keuls tests were performed on RL and SWAT Ratings to determine which means were statistically different (see Table D4). In general, participants responded faster and gave lower SWAT Ratings as they went through more replications of each experimental condition. However, this appeared to stabilize after Block seven. The mean values for Blocks seven through ten are not statistically different for both measures. No two-way or higher-order interaction effect that include Block as one of its variables is significant beyond the .05 level for all three dependent measures. This demonstrated the absence of differential learning across all combinations of Stimulus Class, MSET Size, and Probe Type.

Two ANOVAs for RL and SWAT Ratings were run using the data from Blocks seven through ten only (see Tables D5 and D6). No main or interaction effect involving the Block factor is significant beyond the .06 level. These data, therefore, represent a more stable level of performance than the data from Blocks three through ten. All subsequent analyses include data only from Blocks seven through ten in order to derive a more accurate assessment of the experimental effects.

Primary task performance. A five x four x two ANOVA (Stimulus Class by MSET Size by Probe Type) was performed to assess differences in the amount of secondary task intrusion arising from the various experimental conditions. The data for these ANOVAs were the primary-task RMS tracking performance computed using Equation 1. The results are shown in Table D7. No statistically significant degree of differences in RMS tracking performance across any experimental condition was found.

Secondary task performance. A five x four x two ANOVA (Stimulus Class by MSET Size by Probe Type) was performed on the Sternberg task RL (see Table D8). All three main effects are significant: Stimulus Class [$F(4,36) = 16.64$, $p < .0001$], MSET Size [$F(3,27) = 8.62$, $p = .0004$], and Probe Type [$F(1,9) = 4.44$, $p = .0004$]. Post hoc Newman-Keuls tests were performed on the Stimulus Class and MSET Size effects to determine which levels were statistically different. Tables D9 and D10 present the results of these analyses. Average RL ranged from 780 ms to 997 ms with Digits and Shapes being responded to the quickest and slowest, respectively. The only means not statistically different are between Digits and Words.

The Stimulus Class by MSET Size effect is significant [$F(12,108) = 4.12$, $p < .0001$]. Figure 9 graphically presents this effect. Simple-effects F-tests and Newman-Keuls tests were performed on the MSET Sizes with Stimulus Class held constant. For the Classes of Digits, Words, and Letters, no MSET Size means were statistically different from one another.

Least-squares linear regressions were fitted to each Stimulus Class at the four MSET Sizes. Figure 10 shows these fits and the regression equations. All pairs of slopes (10 pairs for five Stimulus Class slopes) were tested for statistical significance using a method described by Neter and Wasserman (1974, p. 166). The following equation was used:

$$p_1 - p_2 = (b_1 - b_2) \pm t(1 - \propto/2;\ n_1 + n_2 - 4)s(b_1 - b_2) \quad , \tag{2}$$

where: $p_1 - p_2$ refer to the upper and lower confidence limits;

b_1 and b_2 are the slopes to be tested;

$\propto$ is the cut-off point for statistical significance;

n_1 and n_2 are the sample sizes; and

s is the combined sums of squares for the two slopes.

This test determines the statistical difference between two slopes (H_0: $b_1 = b_2$; H_1: $b_1 \neq b_2$).

Results of the slope tests are presented in Table D12. The comparison of any two Stimulus Class slopes whose confidence limits did not involve a sign change (i.e., the range did not span zero) was statistically significant. Slopes for the following Class pairs are significant: Digits-Colors, Digits-Shapes, Colors-Words, Words-Shapes. For these comparisons, it can be said with 95% confidence that as MSET Size increase, RL increase at a statistically faster rate for one Stimulus Class compared to the other.

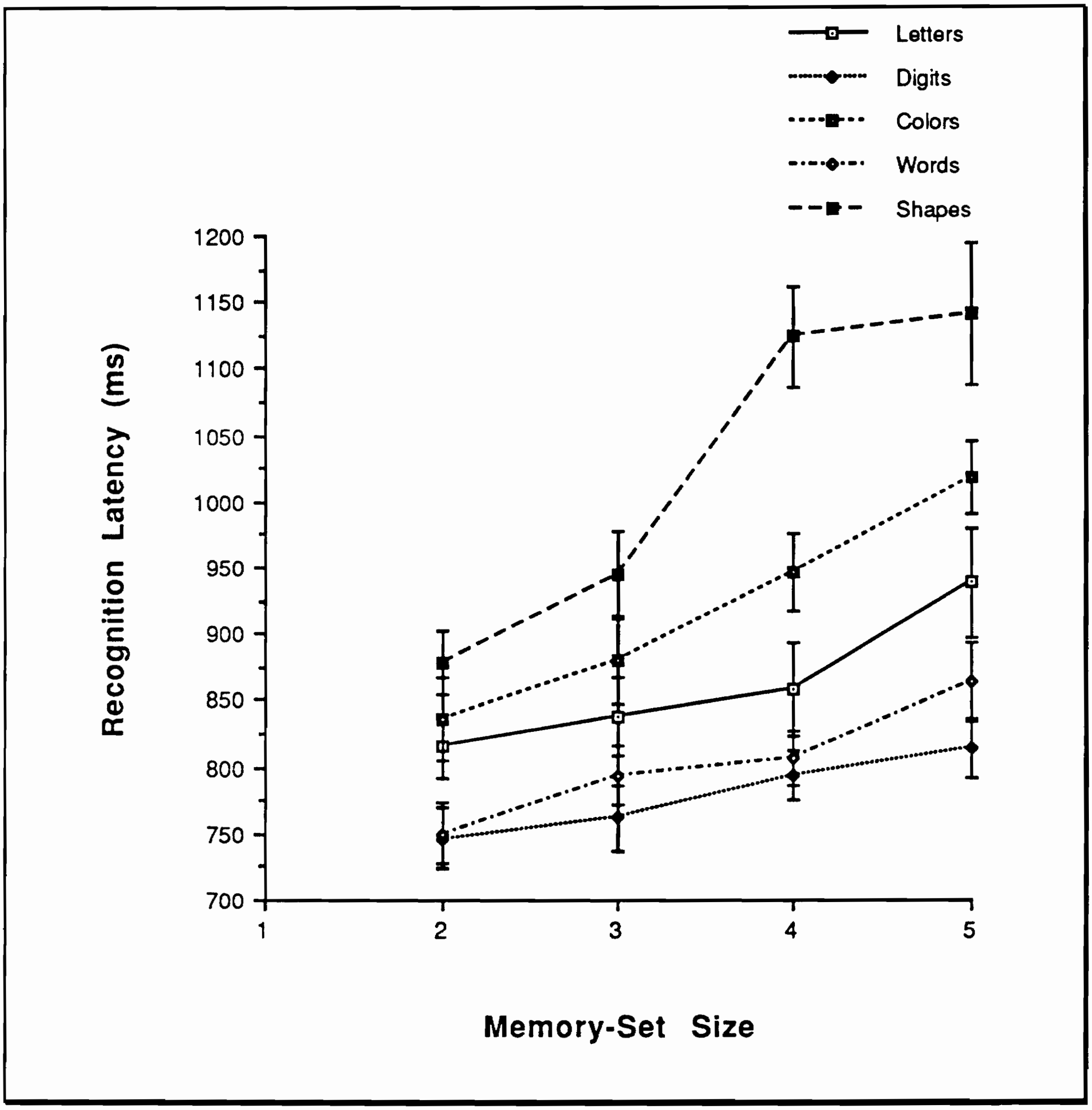


Figure 9. Experiment 1: Secondary Task Recognition Latency as a function of Stimulus Class by Memory-Set Size ($p < .0001$). Data collapsed over Probe Type with N = 10. ± One Standard Error Bars shown.

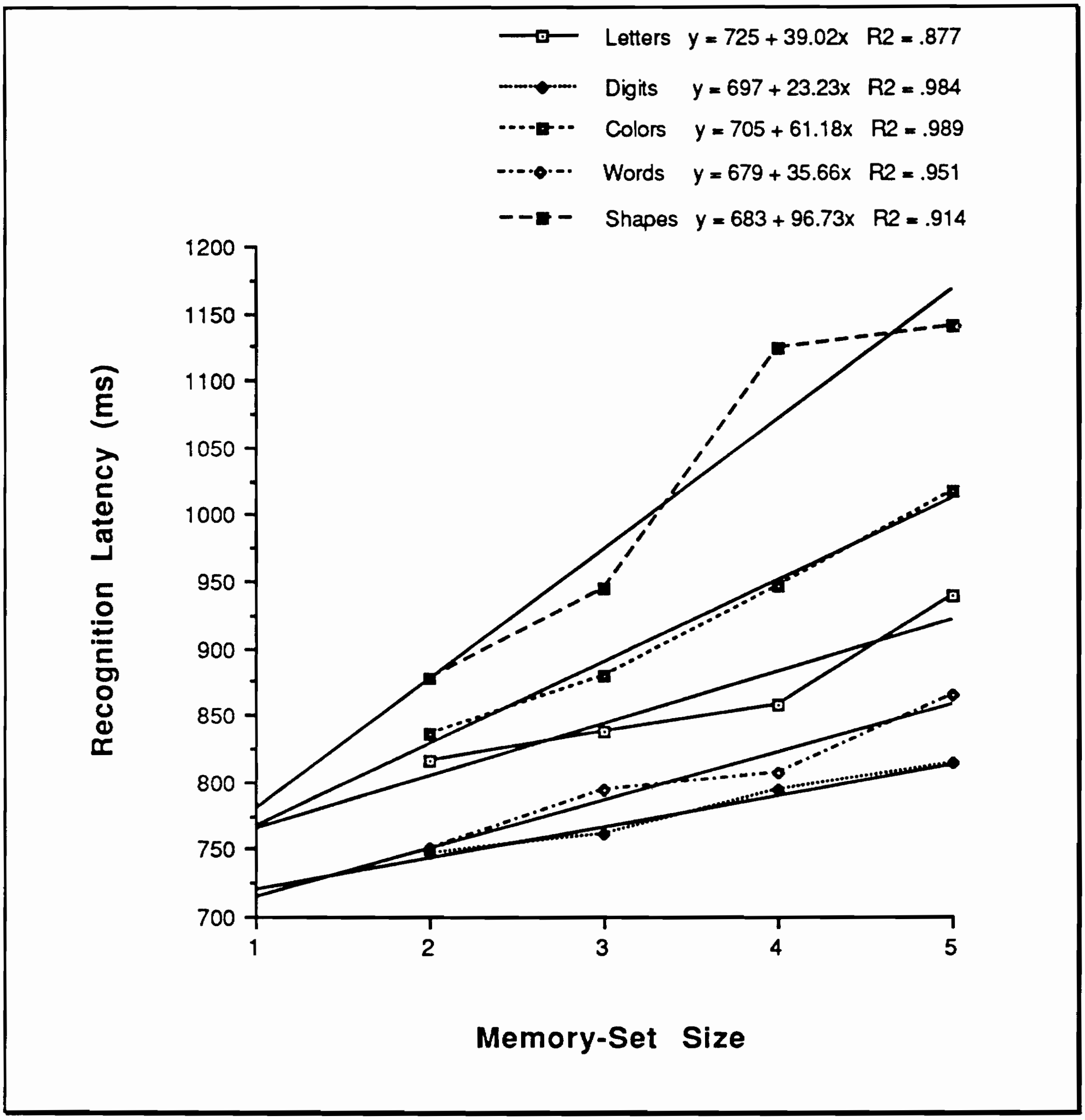


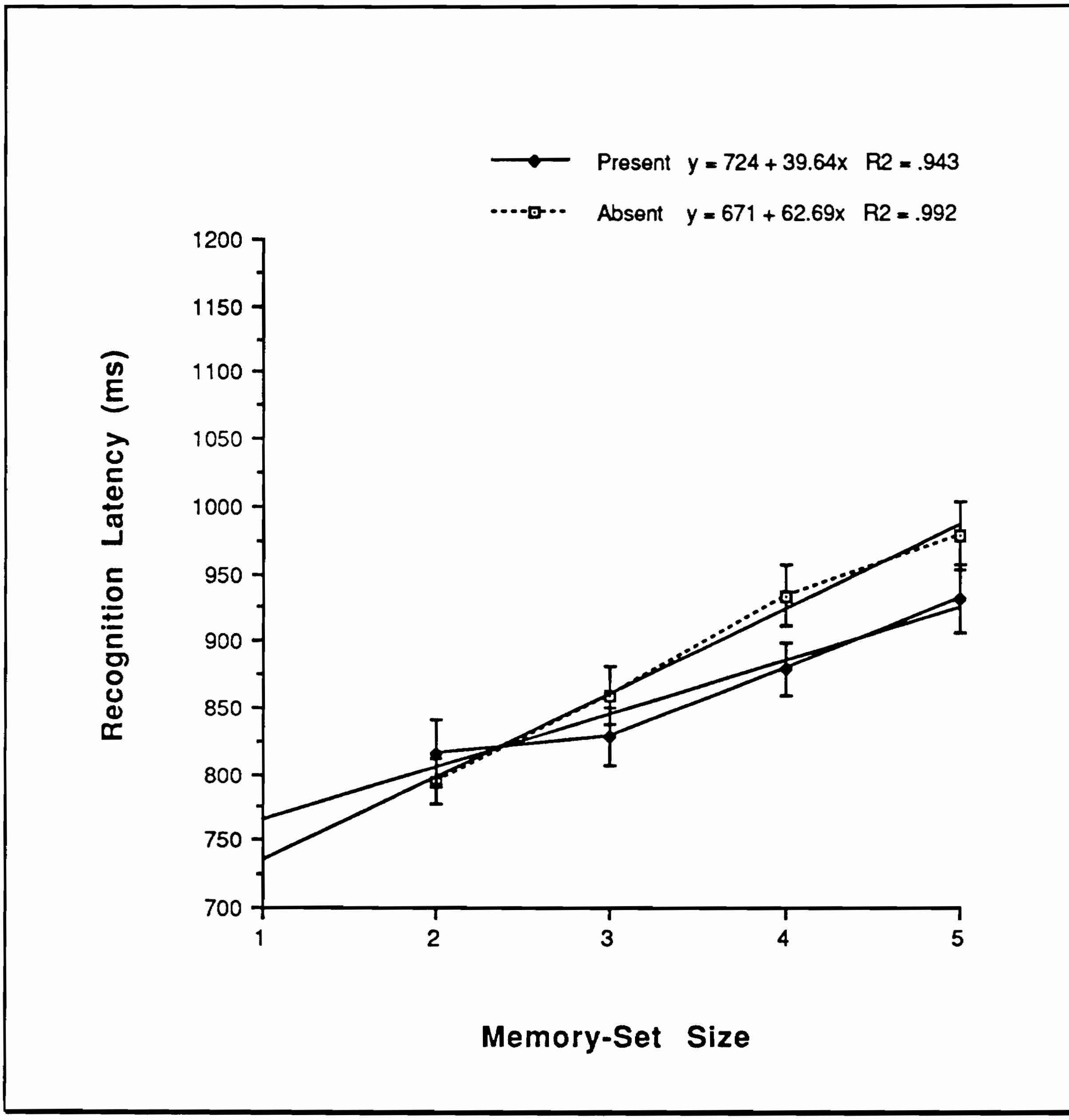
Figure 10. Experiment 1: Least-Squares Linear Regressions fitted to the five Stimulus Classes. 320 data points were used in each Regression.

The MSET Size by Probe Type effect on RL is significant [$F(3,27) = 3.41, p = .0315$]. This effect is graphically portrayed in Figure 11. As Figure 11 shows, negative (i.e., absent) probes elicited longer average RLs than positive or present probes, except for the MSET Size of two. Also shown in Figure 11 is the test of different slopes for present and absent probes (Equation 2 was used for this test). The absence of a sign change across the confidence limits indicates that the slopes are statistically different. As MSET Size increases, RL increases at a faster rate for absent probes than for present probes -- this concurs with the statistically significant MSET Size by Probe Type effect finding.

The average RE across all experimental conditions was 3.375%. Results of a three-way ANOVA (Stimulus Class by MSET Size by Probe Type) performed on this dependent variable are shown in Table D13. The Stimulus Class main effect is significant at $F(4,36) = 2.21, p = .0868$. Figure 12 graphically portrays this effect. Word trials incurred the fewest incorrect responses while Shape trials incurred the most. Comparing Table D9 to Figure 12, the obtained relation of increasing RE across Stimulus Class parallels that for RL (i.e., Classes with longer average RL also incurred more incorrect responses), except for the Digits/Words reversal. Participants responded quicker to Digits than to Words, but made comparatively more errors on Digit trials.

The three-way Stimulus Class by MSET Size by Probe Type effect on RE is significant [$F(12,108) = 1.84, p = .0507$]. This effect is graphically portrayed in Figure 13. The figure should be interpreted with caution because several of the average data points reflect REs at or very close to zero.

Subjective ratings. An analysis of the card sort data was performed to accomplish two objectives. First, the conjoint analysis algorithms assessed the validity of a group scale by performing a series of axiom tests on the rankings. The results of this analysis are shown



-45.1139 ≤ Present - Absent ≤ -0.9761*

Figure 11. Experiment 1: Secondary Task Recognition Latency as a function of Memory-Set Size by Probe Type ($p = .0315$). Data collapsed over Stimulus Class with N = 10. ± One Standard Error Bars shown. Graph also shows Least-Squares Linear Regression fitted to Present and Absent Probes. 800 data points were used in each Regression.

---

* The end points represent confidence limits and an α value of .05 was used for the test (see Equation 2). No sign change indicates statistically different slopes.

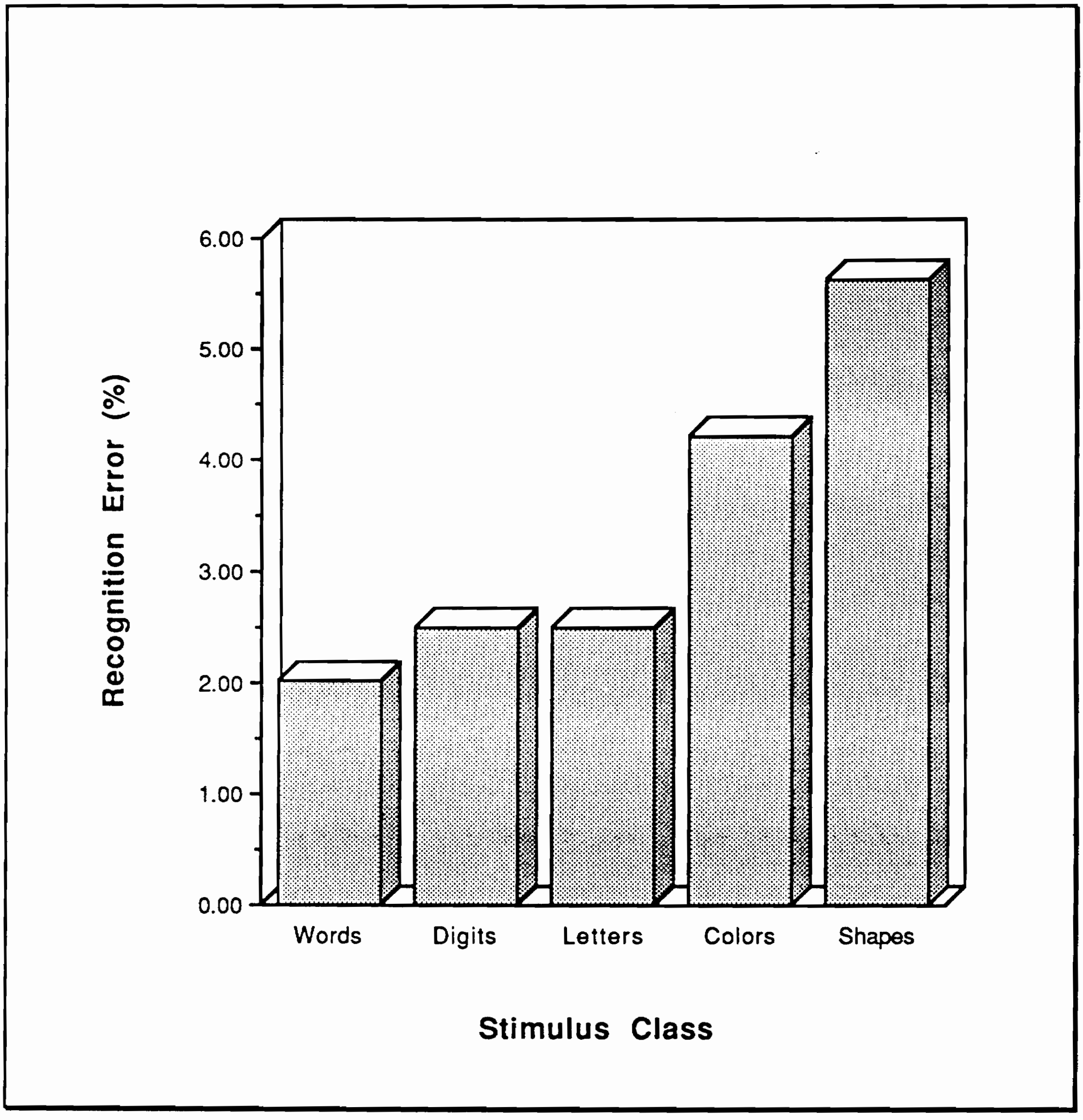


Figure 12. Experiment 1: Secondary Task Recognition Error as a function of Stimulus Class (p = .0868). Data collapsed over Memory-Set Size and Probe Type with N = 10.

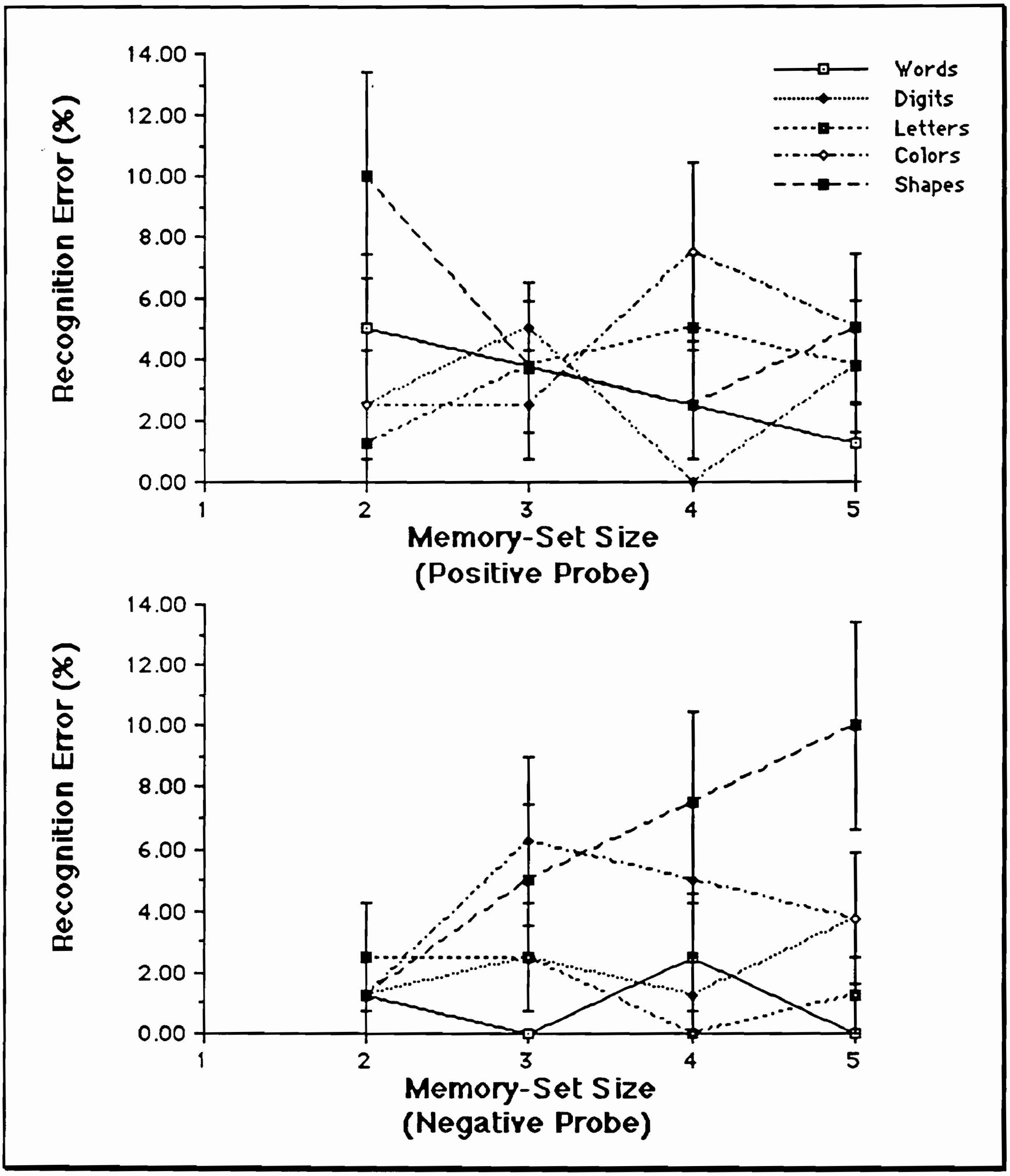


Figure 13. Experiment 1: Secondary Task Recognition Error as a function of Stimulus Class by Memory-Set Size by Probe Type (p = .0507). Data collapsed over N of 10. ± One Standard Error Bars shown.

in Table 2. According to Reid et al. (1989), a Kendall's W of 0.75 or more indicates a high degree of agreement among the raters. The obtained Kendall's W was 0.8023. Reid et al. suggested a criterion of 20 or more axiom test failures before considering a re-sort of the 27 SWAT ranks. Analysis of participants' rankings in this experiment showed a total of 13 axiom test failures. Thus, a single group scale was used for subsequent SWAT analyses. The second objective of SWAT analysis was to compute interval-level rescaled data. This was accomplished using the software supplied by Reid et al. The rescaled values are given in Table 2.

A five x four x two ANOVA (Stimulus Class x MSET Size x Probe Type) was performed on the rescaled, interval-level SWAT Ratings (see Table D14). Two main effects are significant: Stimulus Class [$F(4,36) = 13.58, p < .0001$] and MSET Size [$F(3,27) = 41.07, p < .0001$]. Post hoc Newman-Keuls tests were performed on both effects to determine which levels are statistically different. Tables D15 and D16 show the results of these analyses and the corresponding average SWAT Ratings. For the Stimulus Class effect, mean SWAT Ratings range from 18.62 to 31.00. Only the ratings between Colors and Words are not statistically different. For the MSET Size effect, mean SWAT Ratings range from 6.06 to 43.70, and all means were statistically different from one another.

The Stimulus Class by MSET Size effect also is significant [$F(12,108) = 3.19, p = .0006$; see Figure 14]. Simple-effects F-tests and Newman-Keuls tests were performed on the MSET Sizes with Stimulus Class held constant. Table D17 presents the results of these tests for all five Stimulus Classes. The MSET Size effects are statistically significant at $p < .0001$ for all Classes. All the MSET Size means for all Stimulus Classes are statistically different from one another.

TABLE 2

Experiment 1: Results of SWAT Conjoint Analysis and Interval-Level Rescaled Values for a Group Solution

Kendall's W	0.8023
Axiom Violations	
Independence	0
Double Cancellation	6
Joint Independence	7

Time	Effort	Stress	Rescaled Values
1	1	1	0.0
1	1	2	18.7
1	1	3	37.3
1	2	1	13.9
1	2	2	32.6
1	2	3	51.3
1	3	1	25.2
1	3	2	43.9
1	3	3	62.5
2	1	1	17.7
2	1	2	36.4
2	1	3	55.0
2	2	1	31.6
2	2	2	50.3
2	2	3	68.9
2	3	1	42.8
2	3	2	61.5
2	3	3	80.2
3	1	1	37.5
3	1	2	56.2
3	1	3	74.8
3	2	1	51.4
3	2	2	70.1
3	2	3	88.8
3	3	1	62.7
3	3	2	81.4
3	3	3	100.0

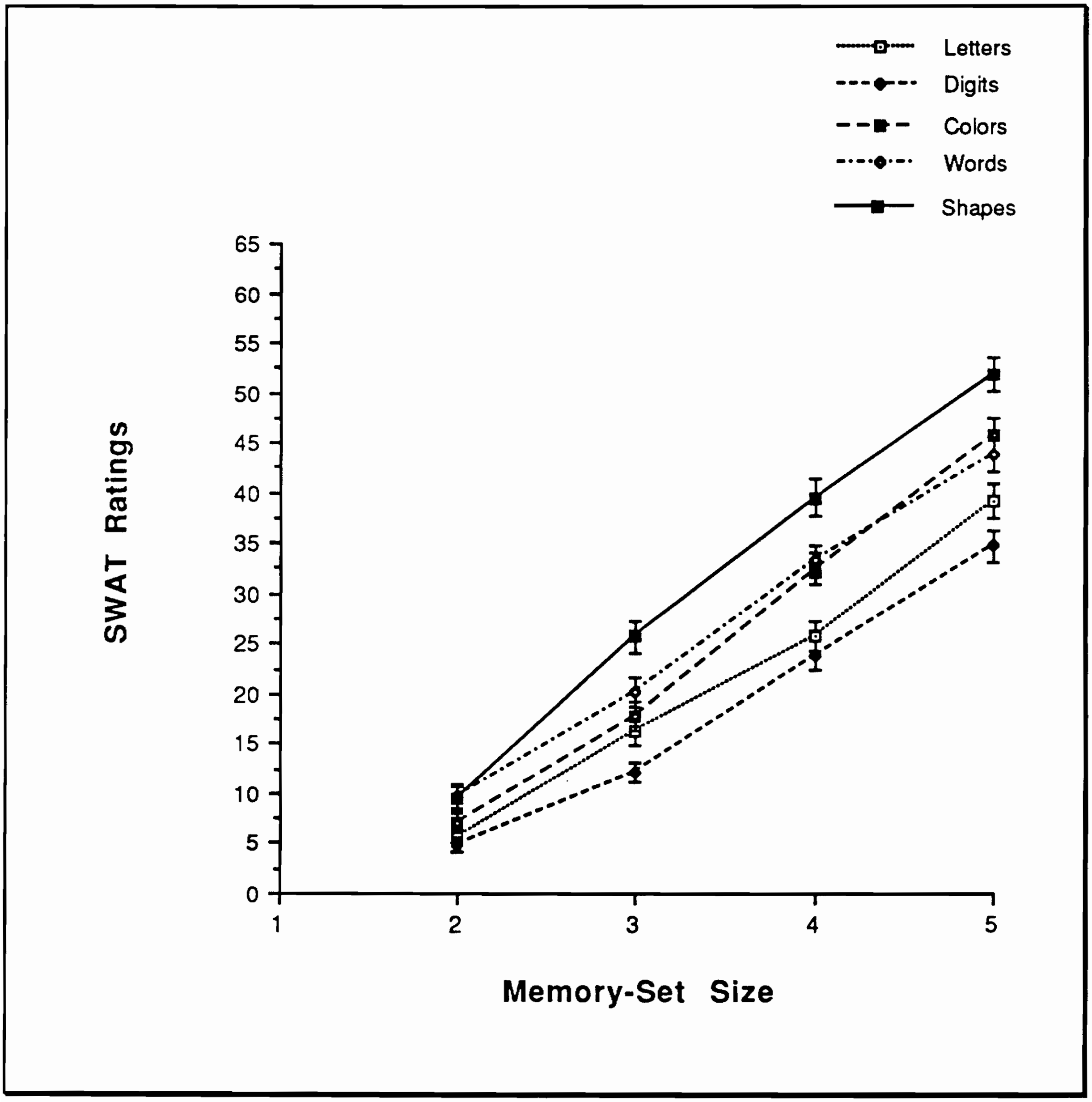


Figure 14. Experiment 1: Secondary Task SWAT Ratings as a function of Stimulus Class by Memory-Set Size (p = .0006). Data collapsed over Probe Type with N = 10. ± One Standard Error Bars shown.

Least-squares linear regressions were computed for each Stimulus Class across the MSET Sizes. The obtained slope coefficients of the regression equations were then pair-wise compared for statistical differences using Equation 2. In all, there were 10 pair-wise comparisons for the five Stimulus Class slopes. The results of these comparisons are presented in Table D18. For any comparison whose confidence limits did not span zero, the slopes are significantly different. The following slope pairs are significantly different: Letters-Shapes, Digits-Colors, Digits-Shapes, Words-Shapes.

The degree of association among SWAT Ratings, RL, and MSET Size was calculated using the Pearson Product Moment correlation procedure. Results of these tests are shown in Table D19. The highest correlation is between SWAT Ratings and MSET Size (coefficient of 0.6757), while the lowest correlation is between RL and MSET Size (coefficient of 0.2358). RL and SWAT Ratings correlate at 0.3614. All correlation coefficients are significant at $p < .0001$.

Discussion

Experiment 1 quantitatively and qualitatively compared memory loading as a function of stimulus class. The Sternberg paradigm provided quantitative performance measures of recognition latency and error, while SWAT provided a qualitative measure of cognitive workload. The following relationship for rate of increase in processing time -- which is also the MSET size slope for RL -- was obtained: digits < words < letters < colors < geometrical shapes (see Figure 15). The relationship for rate of increase in SWAT ratings -- which is the MSET size slope for SWAT ratings -- was: digits < letters < words < colors < geometrical shapes (see Figure 16). The differences shown in these figures are unconfounded due to an absence of differential practice effects across all combinations of stimulus class, MSET size, and probe type. They are, therefore, stable.

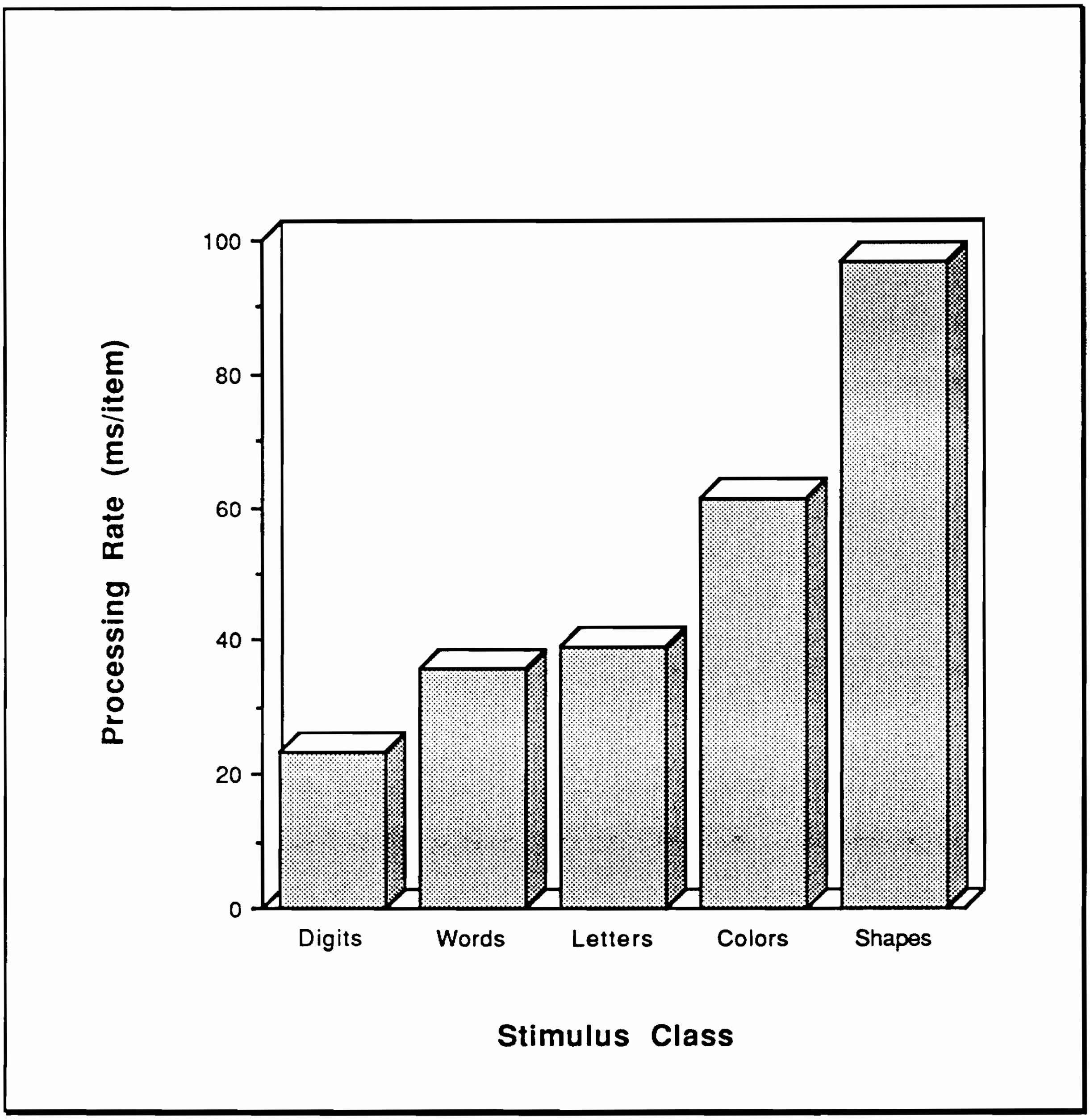


Figure 15. Experiment 1: Stimulus Class processing rate.

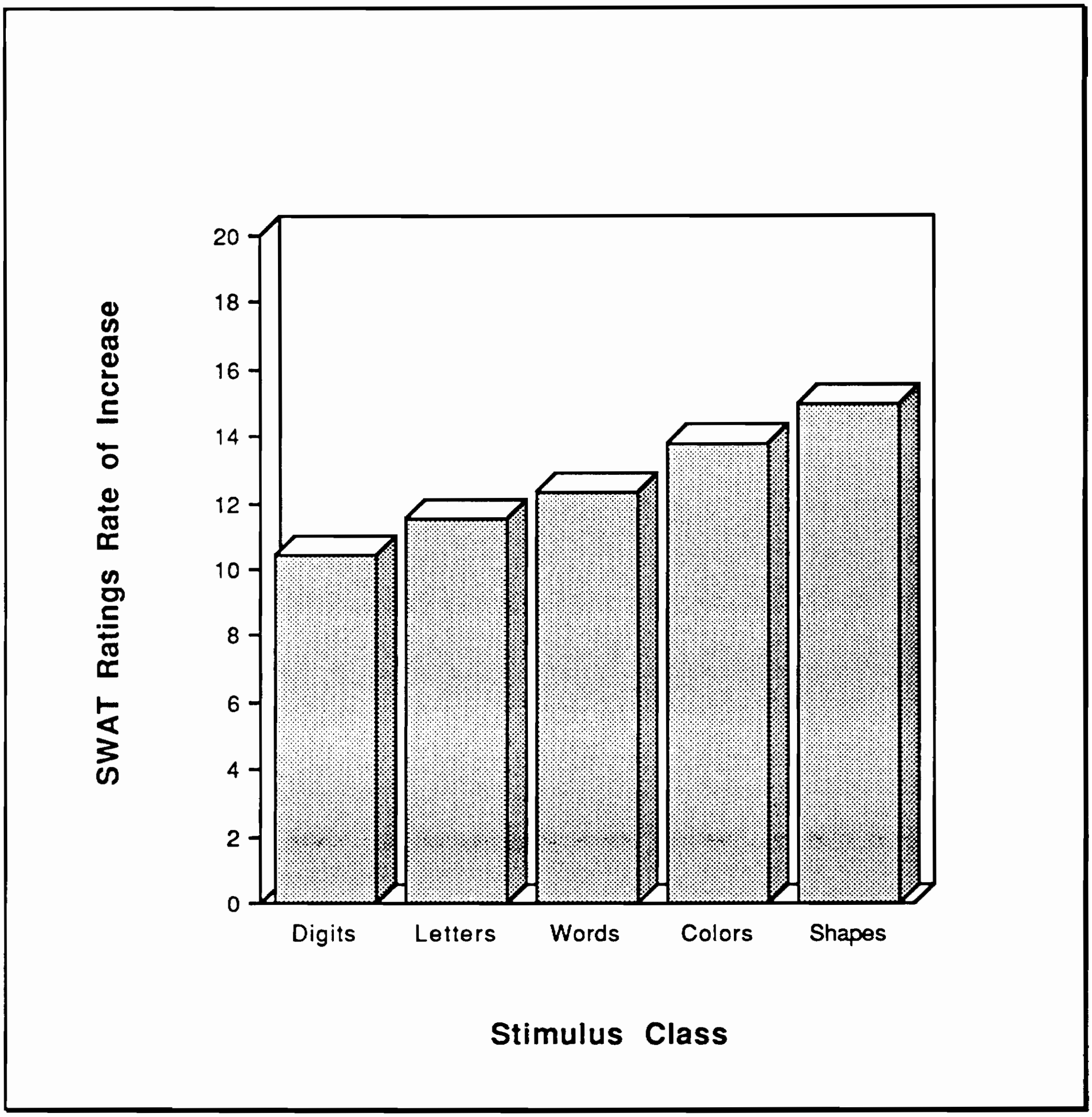


Figure 16. Experiment 1: Stimulus Class SWAT Ratings rate of increase.

Referring to Tables D9 and D10, RL differences among stimulus classes and MSET sizes were considerable in some cases (217 ms difference between digits and shapes; 163 ms difference between MSET sizes two and five). Within each stimulus class, differences among MSET sizes were larger for the classes with longer average RLs (see Table D11). The crucial stimulus class RL differences may be seen best through Figure 10 and Table D12. Both portray differential increases in RL across stimulus class for each additional item held in STM. The largest differences were obtained for the following class pairs: digits and colors, digits and shapes, colors and words, words and shapes. The interpretation is that these class pairs, especially, merit consideration when optimizing stimulus class as an information portrayal variable.

Results from the subjective data demonstrate SWAT to be a sensitive tool for assessing differential memory loading across stimulus class and MSET size (see Tables D15 and D16), at least in a laboratory environment. A comparison of Tables D11 and D17 reveals that SWAT is, in fact, a more sensitive tool than the RL and RE measures obtained using the Sternberg paradigm. All MSET sizes across all stimulus classes were statistically different in terms of SWAT ratings. As was the case for RL, the largest slope differences in terms of SWAT ratings were obtained between stimulus classes with the smallest and largest processing rates (see Table D18). Again, the interpretation is that these classes, especially, merit consideration when optimizing stimulus class as a display variable.

Note that the objective and subjective data dissociated, or failed to agree, in the case of words, letters, and colors. Participants were faster in responding to words than to letters or colors, but they gave higher SWAT ratings for words. This indicates that word recognition requires more cognitive effort in comparison. This finding illustrates the importance of employing two or more assessment categories in CWL research. At issue is the fact that while reaction time data reveal a great deal about our underlying cognitive processes, they should not overshadow the importance of collecting subjective data also (Yeh and Wickens, 1988). This statement is supported by the finding that RL and SWAT ratings correlated at only 0.2924. Stated differently, the use of either measure to predict the other would be successful at a rate of only 8.55%. RL and SWAT ratings, therefore, reveal different facets of the same phenomena (i.e., STM processes). Given the emphasis on both response time and CWL in aviation operations, crew station display design recommendations should be based on both subjective and objective measures.

Recognition error does not vary significantly across experimental conditions. Because a similar relationship was obtained for stimulus class recognition latency and stimulus class recognition error except for the digits/words reversal (compare Figures 12 and 15), this should not present any complication in terms of recommending one stimulus class over another. The nonsignificant MSET size effect is expected since the sizes used did not exceed the generally accepted 7±2 limit.

Performance on the primary tracking task is not statistically different across any experimental condition (see Table D7). This finding is somewhat surprising considering the unfavorable intrusion effects reported by some researchers (e.g., Eggemeier, 1988; Reinhart et al., 1988). This author expected some intrusion differences at least between the smallest and largest MSET sizes. Two interpretations of this finding seem plausible. First, it may be that participants were, indeed, able to time-share effectively both tasks at the low and high workload levels. If this were true, they must have followed the task instructions very closely (i.e., allocate resources to the secondary task only when primary task performance is satisfactory and do not sacrifice accuracy for speed). Second, perhaps the highest workload level manipulated (i.e., the five shapes condition) did not encroach upon STM or channel capacity. This interpretation is supported by the finding of no differences in recognition error across any experimental condition. In either case, the conclusions drawn from this experiment are not affected.

Lead-in to Experiment 2

Experiment 1 investigated memory loading for the case where items from one stimulus class only were held in STM. Experiment 2 extended the investigation to a task context in which items from two stimulus classes need to be processed. The rationale for this experiment is as follows. Considering Baddeley's (1984) working memory model and Wickens (1980) multiple resource theory, STM resources are devoted to retaining and processing information of either a verbal or a spatial nature. The resources available for processing either information type are thought to be drawn from mutually exclusive pools (i.e., verbal information cannot be processed using spatial resources and vice versa). This issue is relevant to stimulus class optimization if it is the case that different stimulus classes are coded and processed using different resource pools. To elaborate, it was mentioned that, in the Gibsonian tradition, it is not unreasonable to suggest that digits, letters, and words afford a verbal encoding scheme while colors and shapes afford a spatial encoding scheme. Because both resource pools are thought to be simultaneously available, STM-resource usage would be more efficient when a portion of incoming information is coded verbally while the rest is coded spatially. In comparison, the use of solely verbal or solely spatial resources would reflect inefficient resource usage.

Continuing with the differential processing efficiency concept, it is predicted that recognition latency for an MSET size of six should be shorter if the items consist of, for example, three words and three shapes than if the items were solely words or solely shapes. In general, it is hypothesized that recognition latency should be shorter when items from two stimulus classes draw resources from the verbal *and* spatial pools than when they draw resources from the verbal pool only or the spatial pool only. The *differential processing code* hypothesis maintains that digits, letters, and words are processed using verbal resources while colors and shapes are processed using spatial resources.

The results of Experiment 1 are cited in support of the above hypothesis. Specifically, colors and shapes elicited longer average RLs in comparison to digits, words, and letters (refer to Table D9). Because concurrent performance on the primary tracking task required spatial resources, fewer resources were available for processing colors or shapes, but not for processing digits, letters, or words. This lack of resource presumably translated into longer RL. As further evidence, contrast Cavanagh's (1972) and Brown and Kirsner's (1980) results with that of Experiment 1 of this study (compare Figures 2 to 15). The two sets of results do not agree for the classes of colors, letters, and words. One plausible explanation for the discrepancy is the following. Unlike this study, Cavanagh and Brown and Kirsner did not employ a second (tracking) task requiring spatial resources. Their subjects may have, therefore, been able to process colors more quickly than letters or words.

The issue is actually more complicated because, as Experiment 1 demonstrated, not all stimulus classes demand the same amount of verbal (or spatial) resources. Shapes, for instance, elicited the longest RL whether or not a concurrent spatial task was being performed. Note that Cavanagh obtained a processing rate of approximately 55 ms/item (without a concurrent spatial task) whereas this study obtained a processing rate of approximately 95 ms/item (with a concurrent spatial task).

EXPERIMENT 2

Experiment 1 demonstrated the validity of the differential processing efficiency concept for the case where items from one stimulus class were held in STM. Experiment 2 was designed to extend the investigation to the case where items from two stimulus classes are retained and processed in STM. In this case, the differential STM investigation addresses verbal and spatial encoding and processing schemes of different stimulus classes.

Objective, Hypothesis, and Justification

Stimulus class was tested as an information coding method when items from two classes have to be processed in STM. It is hypothesized that some stimulus classes should be more efficient than others in terms of the time and cognitive effort needed to search through a memory set in an item recognition task. Selection of the two-stimulus-class item recognition task is based on Baddeley's (1984) working memory model, Wickens (1980) multiple resource theory, and the results of Experiment 1.

Experimental Tasks

A departure from the task procedure of Experiment 1 was used in this experiment. Specifically, only one main task was used to load STM. An example trial is shown in Figure 17, in which an MSET size of eight includes four letters and four shapes. The items have been randomly positioned inside a 24 cm x 20 cm rectangle. Participants were required to memorize these items and, then, respond to a single probe that may have been included in the MSET. The SWAT scale was used as a subjective measure of mental loading while recognition latency and error served as objective measures.

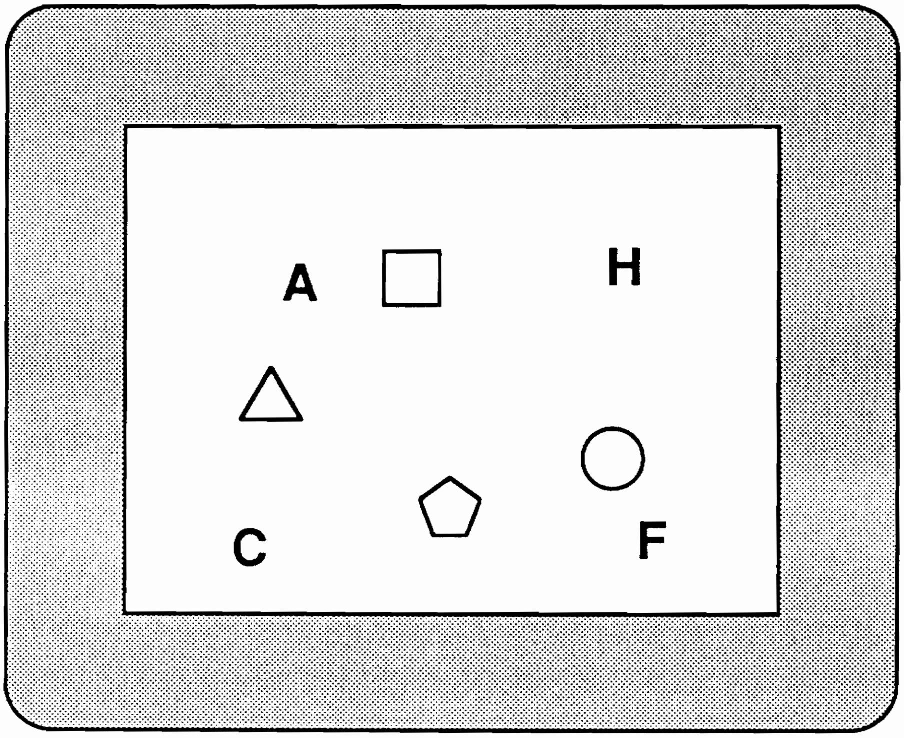


Figure 17. Experiment 2: A representative stimulus display containing two stimulus classes (letters and shapes) of four items each. Items appear serially at a rate of 1.5 s each.

Methodology

Research participants, selection process, apparatus, and stimulus. This experiment employed 10 male individuals, aged between 20 and 27 years, solicited from the student population of Virginia Tech. The participants of this experiment were different from those of Experiment 1. All underwent the same informed consent and screening procedures as in Experiment 1. Appendix A presents a copy of the informed consent contract signed by all participants. All were paid $25 for their time. The same apparatus, stimulus classes, and stimulus class items used in Experiment 1 were employed for the present experiment (see Figure 8 for stimulus class items used).

Design and procedure. The experiment was conducted over five sessions. The first session lasted approximately 2 hr, with the first hour devoted to development of the SWAT rating scale (see card sort instructions in Appendix B), while the second hour was used for administering task instructions and one practice block of the experimental tasks. Actual data were collected on Sessions two through five, which lasted approximately 1 hr each.

MSET items for each trial were selected randomly from two stimulus classes, with two, three, or four items drawn from each class (see Figure 18 for specific class item combinations). As Figure 18 shows, MSET sizes of either four (2 stimulus classes x 2 items from each class), six (2 x 3), or eight (2 x 4) items were presented on any trial. Over trials, the probe item was selected from either of the stimulus classes. For half of the trials, the probe item was present in the MSET (i.e., probe probability was set at 50%).

A typical trial proceeded as follows. Participants initiated the trial by pressing the "YES" key on the response box. Immediately, the MSET items began appearing at a rate of 1.5 s each. After presentation of the last item, there was a 2-s delay and, then, the probe item was shown. Participants pressed either the "YES" or "NO" key to indicate whether

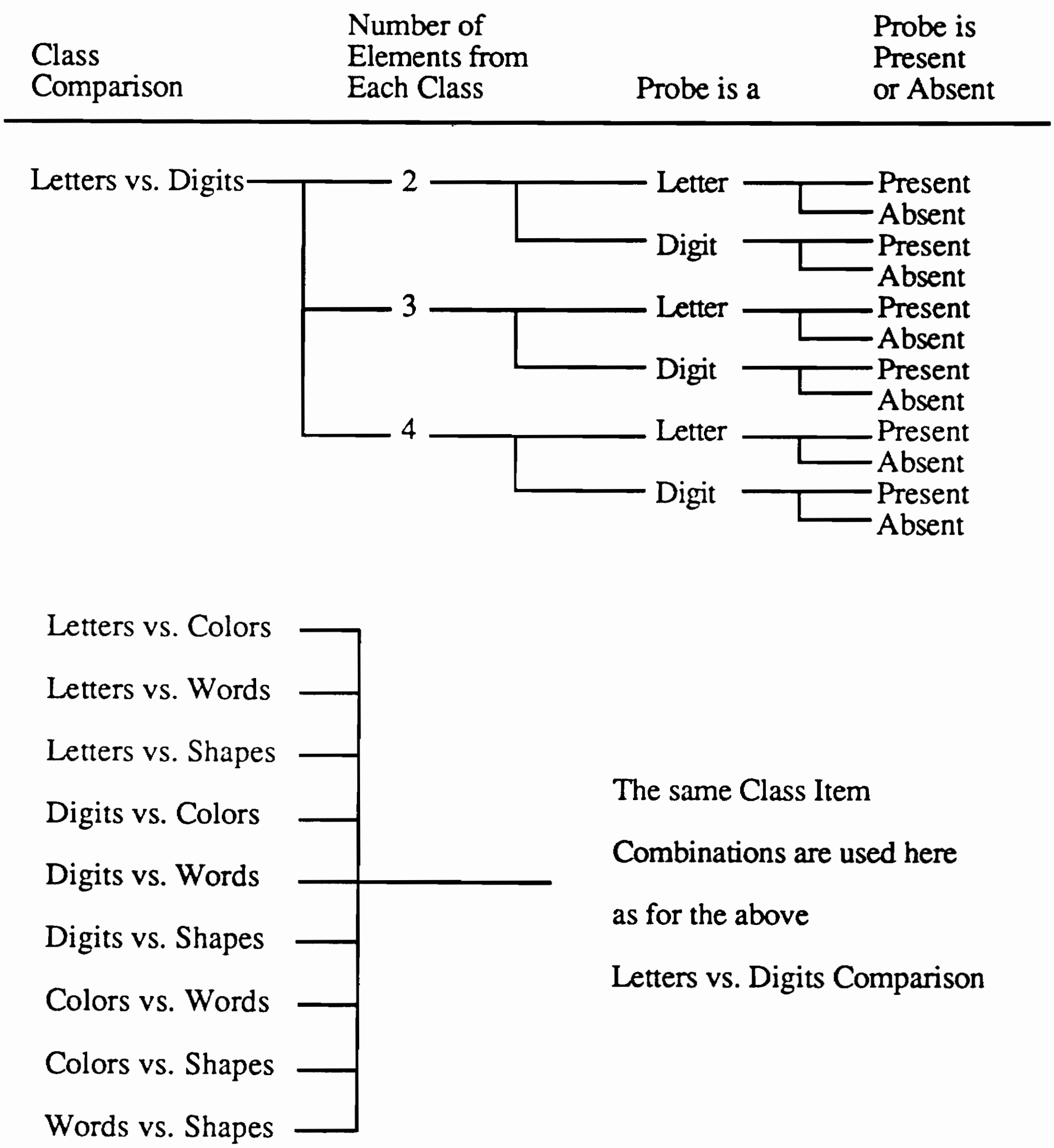


Figure 18. Experiment 2: Class Item combination used in each Class-Comparison.

the probe was part of the MSET. The time from probe onset to response constituted recognition latency (RL). After giving their response, the SWAT dimensions appeared and participants rated the amount of CWL associated with performing the trial. Participant instructions for this experiment are shown in Appendix E.

Figure 18 also shows the 12 conditions that were run for each class-pair comparison. With 10 class-pair comparisons, there was a total of 120 (12 x 10) experimental conditions. Each condition was replicated in a unique random order four times per participant, and each replication was conducted on a different day. Thus, there were 480 (120 x 4) trials for each participant.

Results

Card sort data analyses. As in Experiment 1, the card sort data in this experiment were analyzed to assess their conformance to an additive model and to create interval-level rescaled values for each of the 27 ordinal rankings. Results of these analyses are given in Table 3. The obtained Kendall's W of 0.8834 indicates considerable agreement among participants as to the priority given to the three SWAT dimensions. Moreover, since there was a total of only six axiom test failures, SWAT prototyping was not used. A group scale was developed for the subsequent analyses.

As in Experiment 1, initial data analyses centered on the effects of learning over the four blocks of trials. The Block effect was analyzed along with the factors of Class Comparison, Class Size, and Probe Type using an ANOVA procedure. Tables F1, F2, and F3 show the ANOVA results for recognition latency (RL), recognition error (RE), and SWAT Ratings, respectively. The Block effect is significant for RL [$F(3,27) = 5.39$, $p = .0049$], but not RE [$F(3,27) = 1.99$, $p = .1396$] or SWAT Ratings [$F(3,27) = 2.24$, $p = .1064$]. Results of a Newman-Keuls test performed on the Block effect for RL are shown

TABLE 3

Experiment 2: Results of SWAT Conjoint Analysis and Interval-Level Rescaled Values for a Group Solution

Kendall's W	0.8834
Axiom Violations	
Independence	0
Double Cancellation	0
Joint Independence	6

Time	Effort	Stress	Rescaled Values
1	1	1	0.0
1	1	2	18.3
1	1	3	42.7
1	2	1	18.2
1	2	2	36.5
1	2	3	60.9
1	3	1	27.3
1	3	2	45.7
1	3	3	70.0
2	1	1	11.7
2	1	2	30.0
2	1	3	54.3
2	2	1	29.9
2	2	2	48.2
2	2	3	72.5
2	3	1	39.0
2	3	2	57.3
2	3	3	81.6
3	1	1	30.0
3	1	2	48.4
3	1	3	72.7
3	2	1	48.2
3	2	2	66.6
3	2	3	90.9
3	3	1	57.3
3	3	2	75.7
3	3	3	100.0

in Table F4. The means for Blocks two, three, and four are not statistically different from one another.

The Block by Class Comparison by Class Size effect for RE is significant [$F(54,486) = 1.55$, $p = .0096$; see Table F2]. It is not apparent why significance is obtain for this higher-order effect. Due to the absence of differential learning effects (for the dependent variables of RL and SWAT Ratings), data from all blocks (not including the practice block) were used in the analyses.

Class-Comparison analyses. Because the data from all Class Comparisons could not be collectively analyzed using any known ANOVA procedure (it is neither a factorial nor a nested design), each Class Comparison was analyzed separately. The variables of concern for each Class Comparison are Class Size, Probe Class, and Probe Type. Tables F5 to F14 present ANOVA results for each Class Comparison for the dependent variable RL. Of interest are the Probe Class and Class Size by Probe Class effects. Thus, the Probe Class RL for each Class Comparison is presented in Table F15; also shown is the associated Probe Class effect p-value. The following six Class Comparisons had significant Probe Class effects: Digits-Letters, Digits-Shapes, Words-Letters, Words-Shapes, Letters-Shapes, and Colors-Shapes. These effects at the three MSET sizes are graphically presented in Figures 19 to 24.

RL differences among all 10 Class Comparisons was analyzed using a one-way ANOVA. This effect is significant [$F(9,81) = 7.13$, $p < .0001$]. The results of a Newman-Keuls test performed on this effect are presented in Table F16.

The overall RE was 3.771%. A one-way ANOVA performed across the 10 Class Comparisons is significant [$F(9,81) = 3.29$, $p = .0018$]. The results of a Newman-Keuls

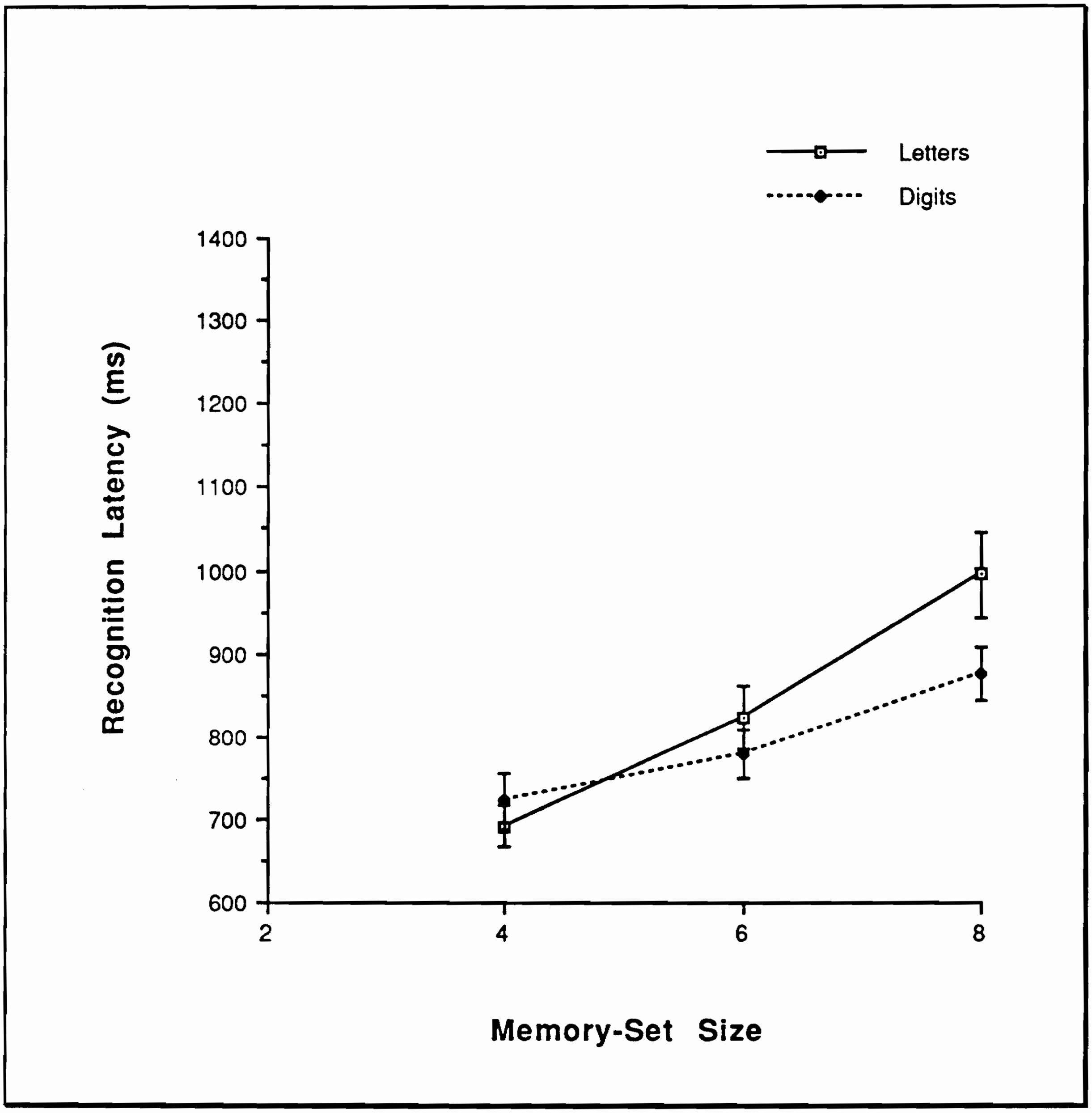


Figure 19. Experiment 2: Recognition Latency as a function of Probe Class by Memory-Set Size for the Letters versus Digits Comparison (p = .2902). Data collapsed over Probe Type with N = 10. ± One Standard Error Bars shown.

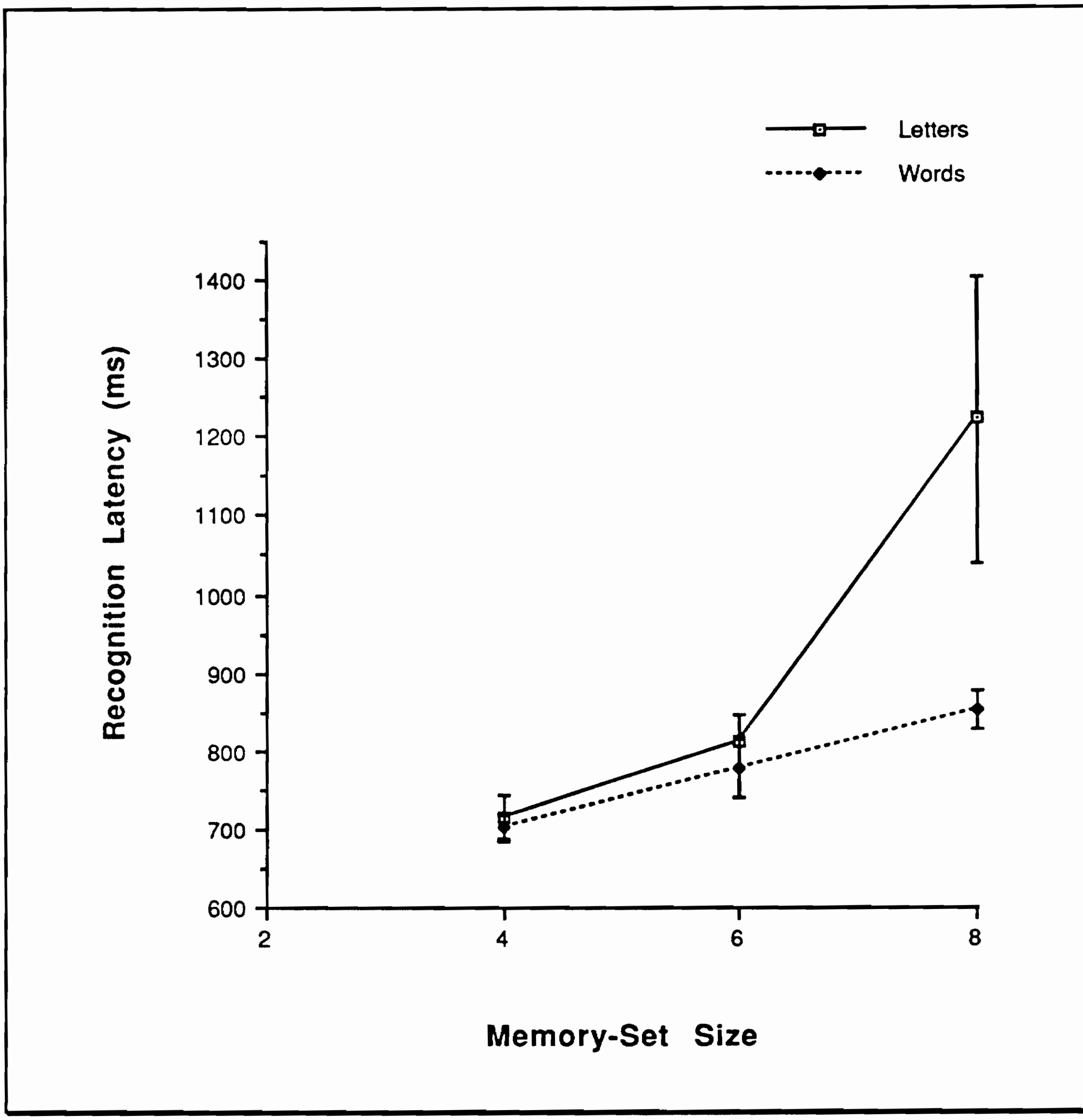


Figure 20. Experiment 2: Recognition Latency as a function of Probe Class by Memory-Set Size for the Letters versus Words Comparison (p = .1297). Data collapsed over Probe Type with N = 10. ± One Standard Error Bars shown.

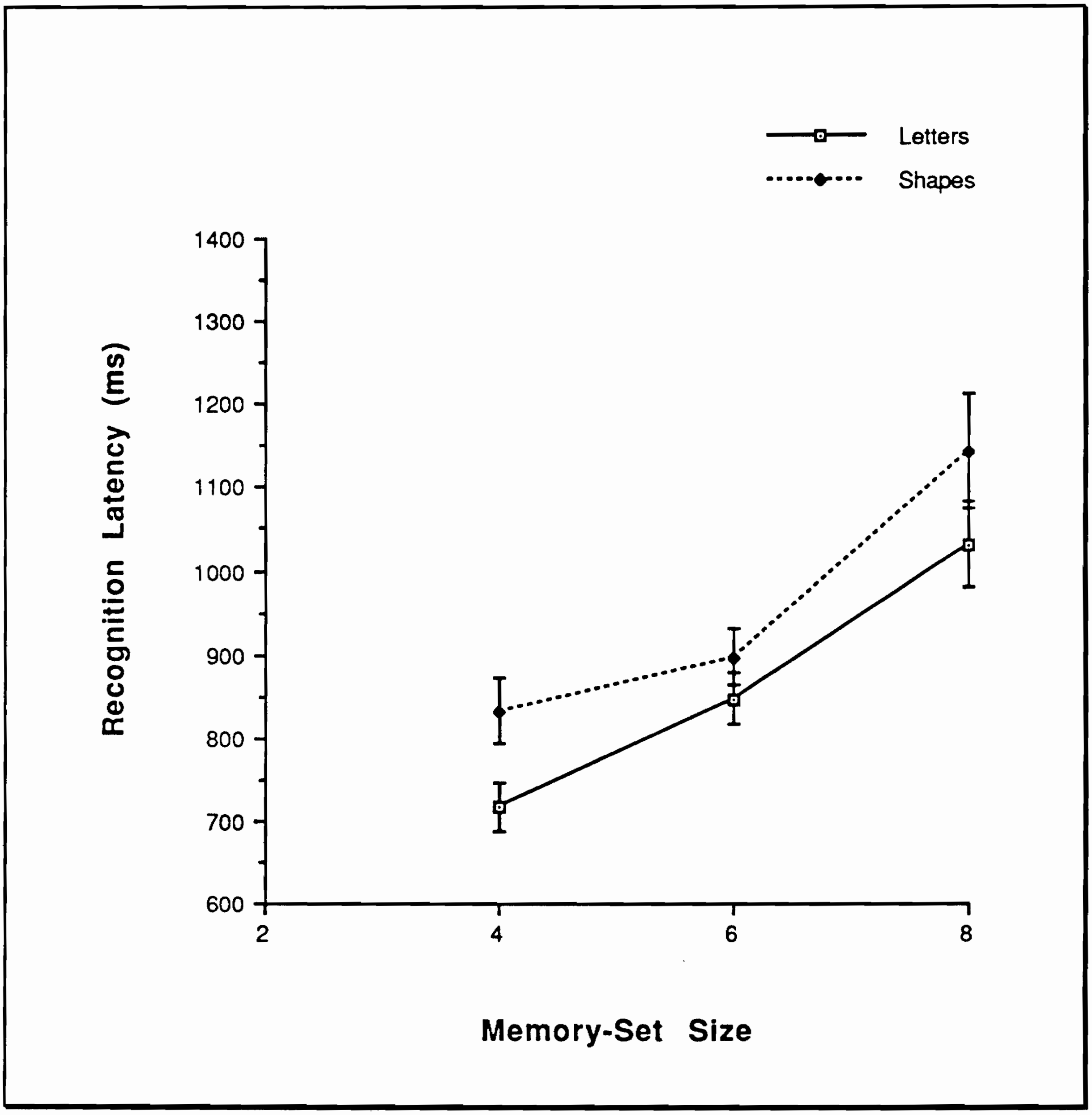


Figure 21. Experiment 2: Recognition Latency as a function of Probe Class by Memory-Set Size for the Letters versus Shapes Comparison ($p = .7891$). Data collapsed over Probe Type with N = 10. ± One Standard Error Bars shown.

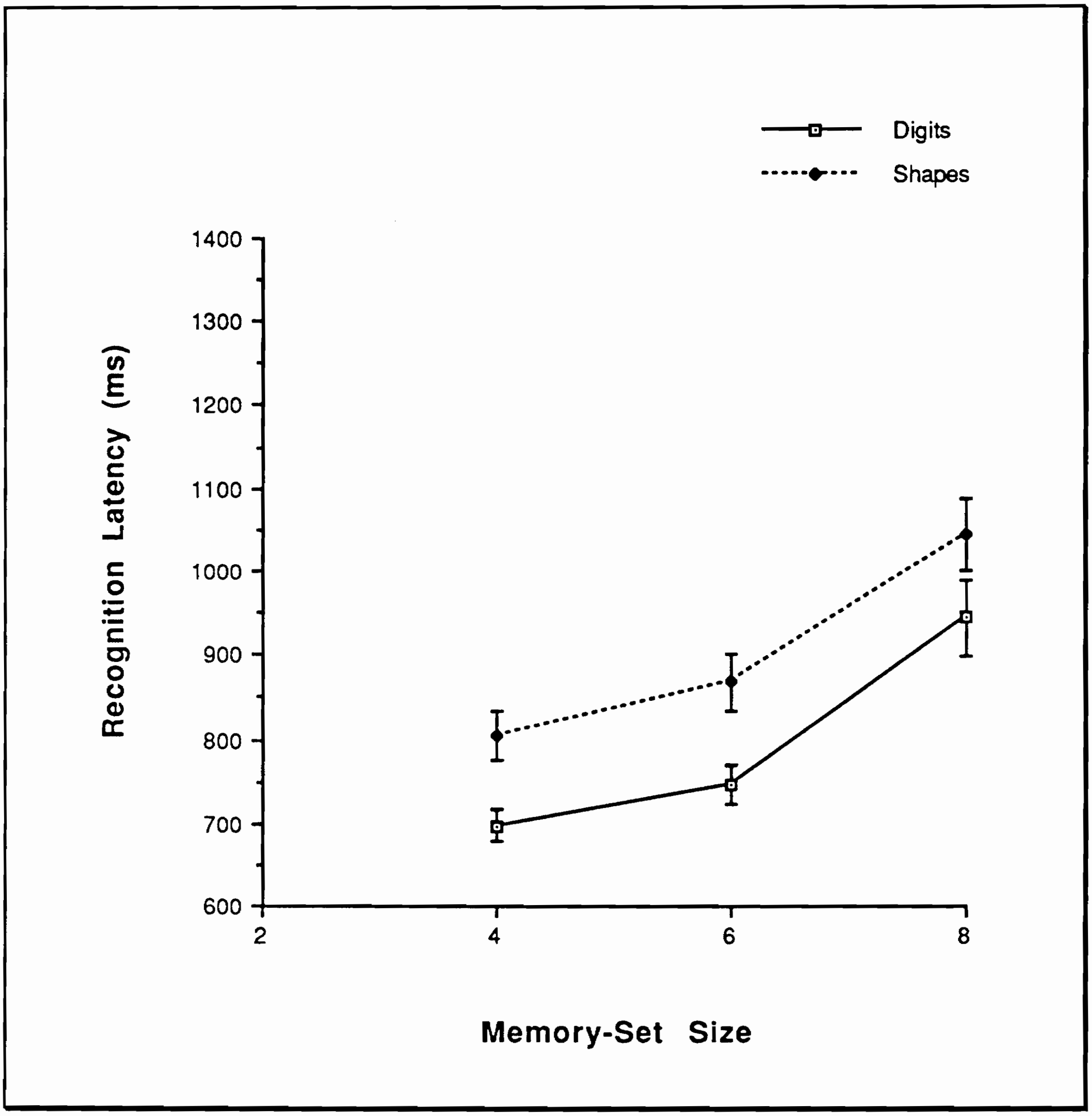


Figure 22. Experiment 2: Recognition Latency as a function of Probe Class by Memory-Set Size for the Digits versus Shapes Comparison (p = .9789). Data collapsed over Probe Type with N = 10. ± One Standard Error Bars shown.

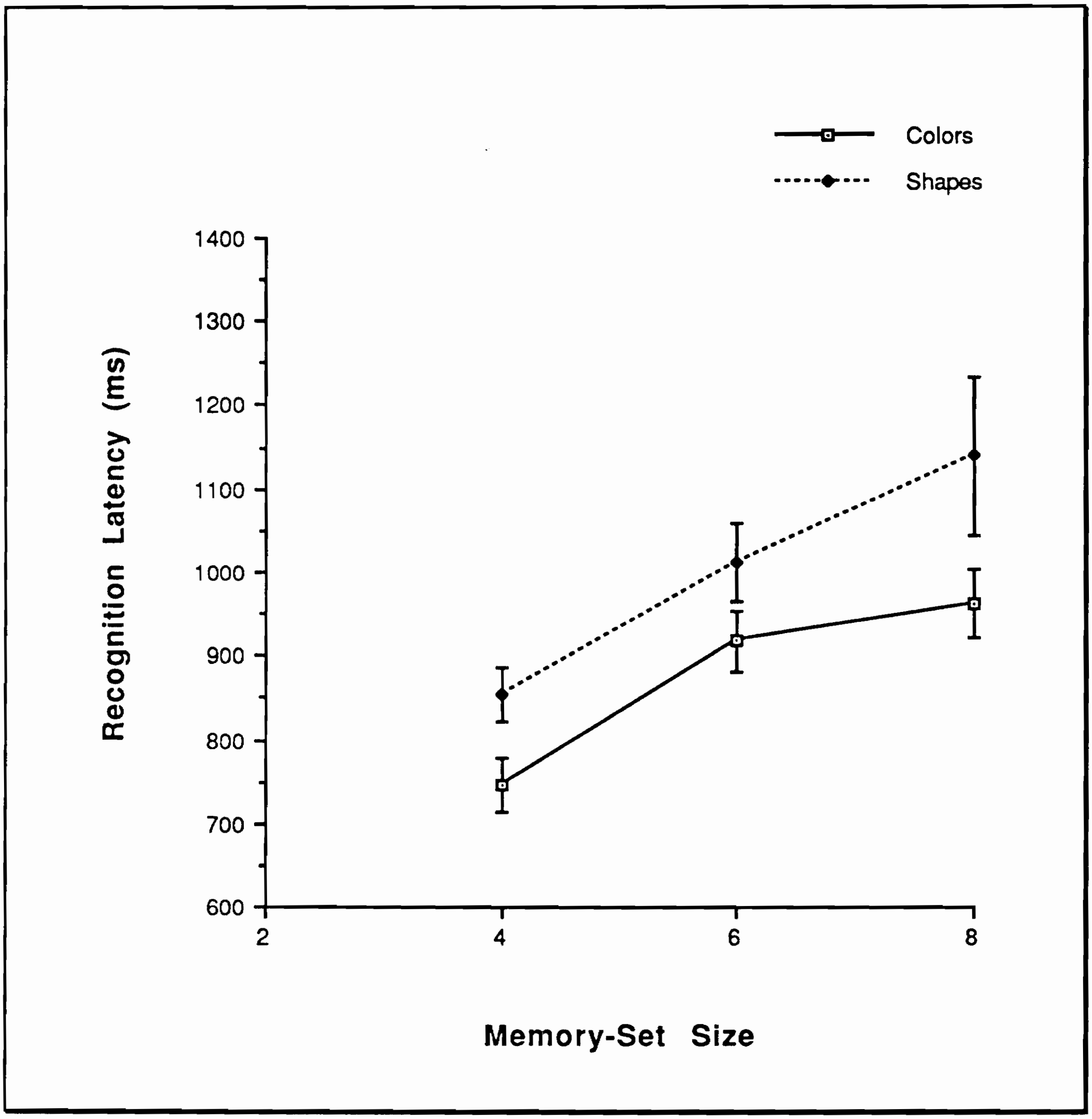


Figure 23. Experiment 2: Recognition Latency as a function of Probe Class by Memory-Set Size for the Colors versus Shapes Comparison (p = .7931). Data collapsed over Probe Type with N = 10. ± One Standard Error Bars shown.

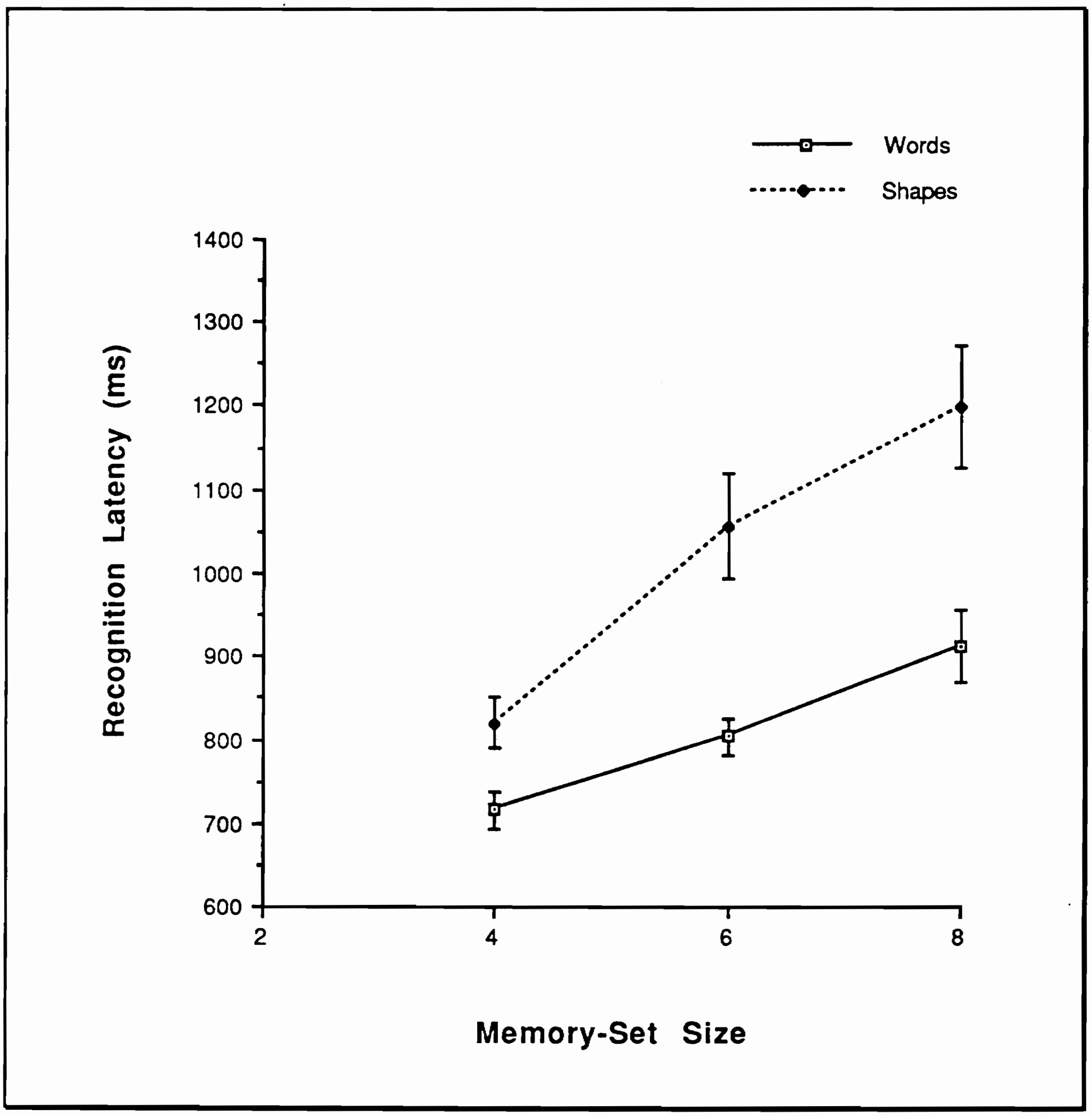


Figure 24. Experiment 2: Recognition Latency as a function of Probe Class by Memory-Set Size for the Words versus Shapes Comparison (p = .0850). Data collapsed over Probe Type with N = 10. ± One Standard Error Bars shown.

test performed on this effect are presented as part of Table F16. In general, Class Comparisons with longer average RLs also incurred more incorrect responses.

Tables F17 to F26 present ANOVA results for each Class Comparison for the dependent variable of SWAT Ratings. Table F27 lists the SWAT Ratings for each Probe Class and the associated Probe Class effect *p*-value for each Class Comparison. Probe Class effects for the following four Class Comparisons are significant: Digits-Shapes, Words-Shapes, Letters-Shapes, and Colors-Shapes. These effects at the three MSET sizes are graphically portrayed in Figures 25 to 28.

A one-way ANOVA performed on SWAT Ratings differences across Class Comparison found this effect to be significant [$F(9,81) = 21.04$, $p < .0001$]. Newman-Keuls test results for this effect are presented as part of Table F16. Average SWAT Ratings range from 16.9 (for the Letters vs. Digits comparison) to 30.9 (for the Colors vs. Shapes comparison).

The relationship among MSET Size, RL, and SWAT Ratings was calculated using the Pearson Product Moment correlation Procedure (see Table F28). The highest correlation was obtained between SWAT Ratings and MSET Size (coefficient of 0.6966), while the lowest was between RL and MSET Size (coefficient of 0.2412). SWAT Ratings and RL correlated with a coefficient of 0.3232. All variable pairs statistically co-varied at $p < .0001$.

Discussion

The results of this experiment generally concur with the findings of Experiment 1. Stimulus class comparisons with large RL and SWAT rating differences resulted in smaller *p*-values for the experimental effects than comparisons with small RL and SWAT rating

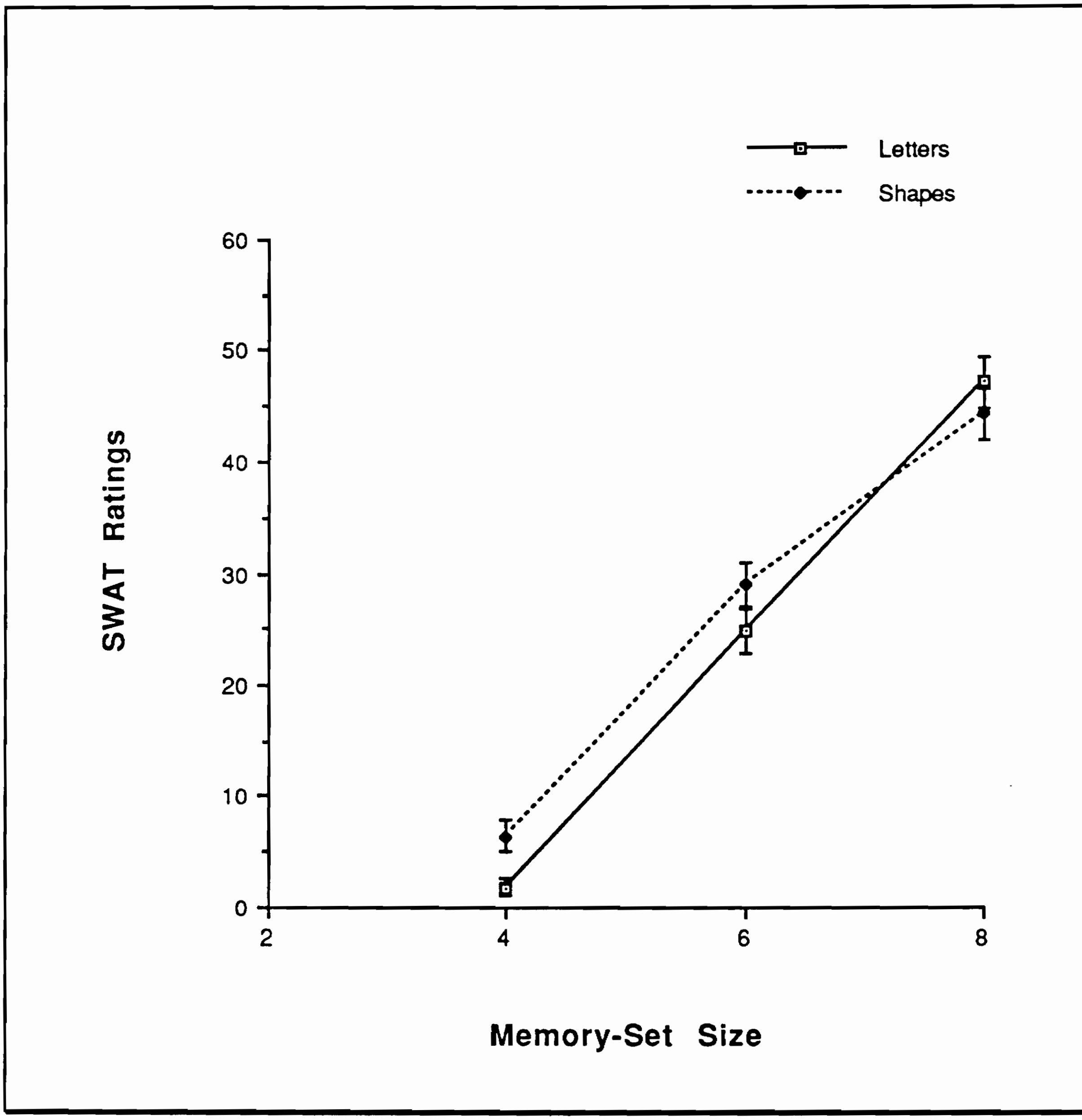


Figure 25. Experiment 2: SWAT Ratings as a function of Probe Class by Memory-Set Size for the Letters versus Shapes Comparison (p = .1137). Data collapsed over Probe Type with N = 10. ± One Standard Error Bars shown.

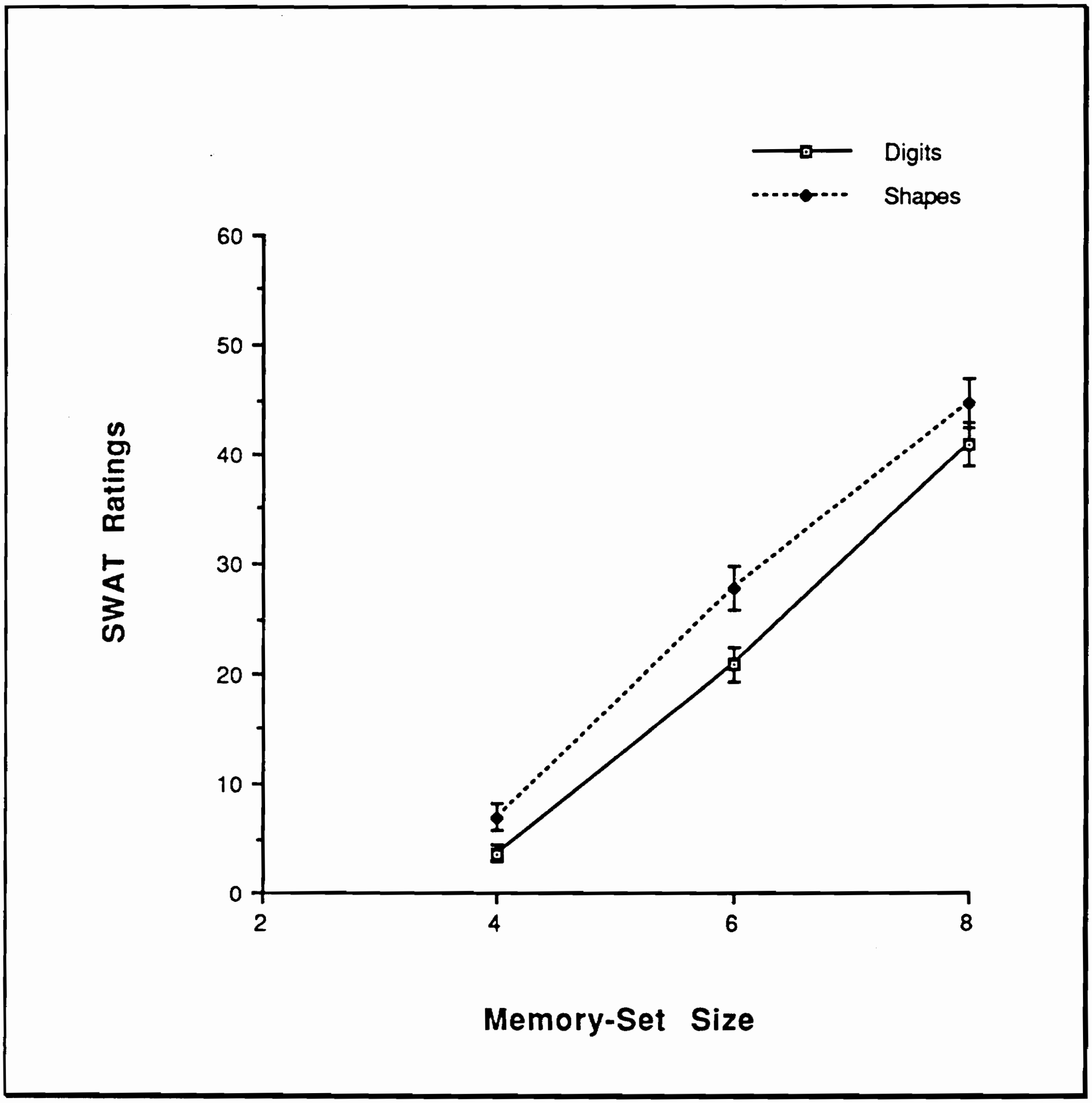


Figure 26. Experiment 2: SWAT Ratings as a function of Probe Class by Memory-Set Size for the Digits versus Shapes Comparison ($p = .2671$). Data collapsed over Probe Type with N = 10. ± One Standard Error Bars shown.

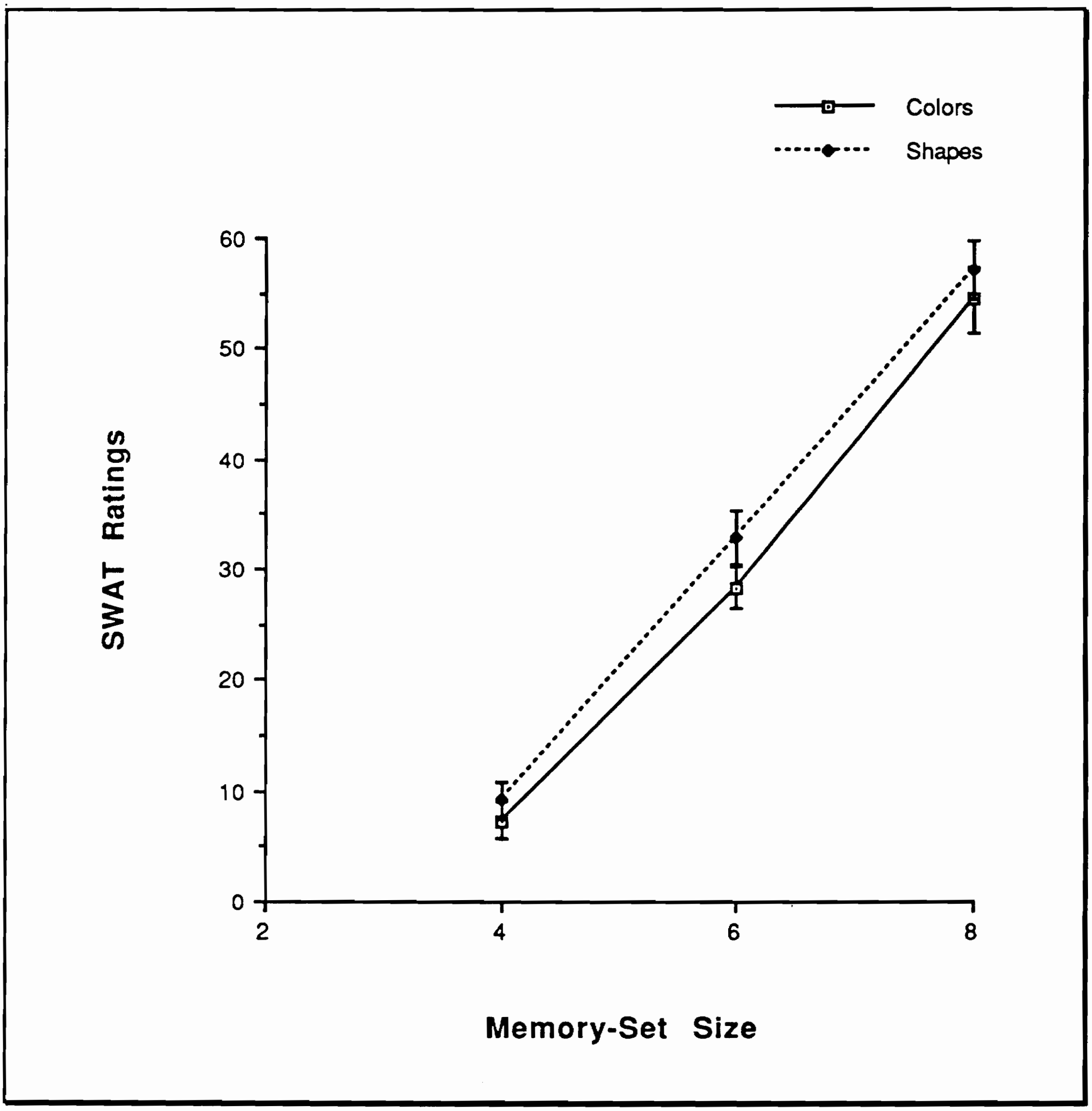


Figure 27. Experiment 2: SWAT Ratings as a function of Probe Class by Memory-Set Size for the Colors versus Shapes Comparison (p = .7192). Data collapsed over Probe Type with N = 10. ± One Standard Error Bars shown.

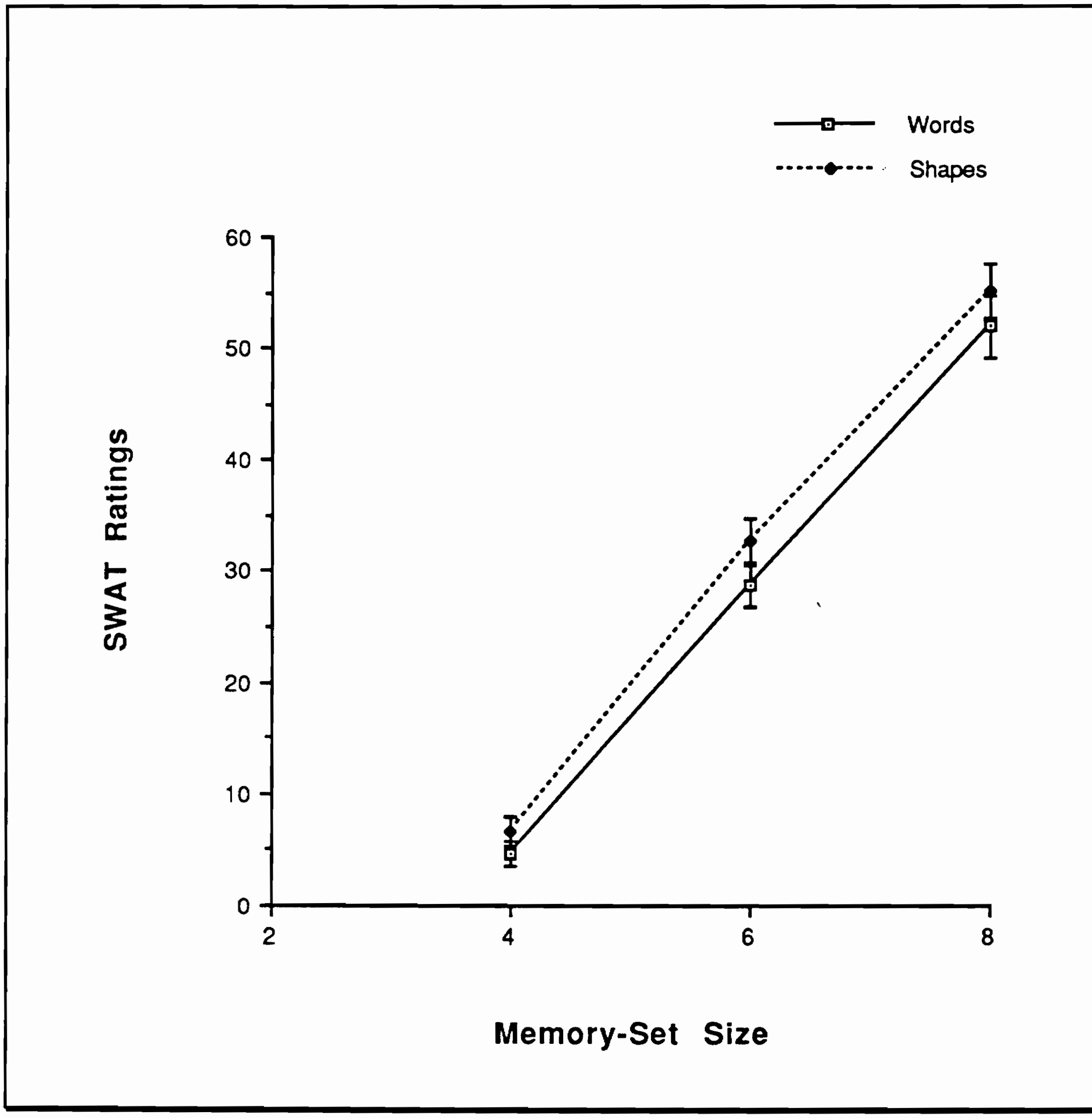


Figure 28. Experiment 2: SWAT Ratings as a function of Probe Class by Memory-Set Size for the Words versus Shapes Comparison (p = .4759). Data collapsed over Probe Type with N = 10. ± One Standard Error Bars shown.

differences. For example, the Probe Class effect for the Digits-Shapes comparison is significant [$F(1,9) = 9.74$, $p = .0123$], while the Probe Class effect for the Digits-Words comparison is not significant [$F(1,9) = 0.67$, $p = .4347$; see Table F15].

Although not as clearly interpretable, a degree of measurement instrument dissociation between RL and SWAT ratings was obtained in this experiment. The Digits-Words comparison, for example, elicited a quicker average RL than the Letters-Digits comparison but was rated higher in cognitive effort (see Table F16). In general, the dissociation is largely confined to the class comparison pairs whose RL and SWAT rating differences are not statistically different.

Table F16 shows that class comparisons with longer RLs also elicited more incorrect responses. In one respect, this indicates that there is one less issue that display designers need to be concerned with (i.e., display designers can focus on the stimulus class combinations that will elicit the shortest recognition times without having to be concerned that these combinations may result in comparatively more recognition errors).

Following the theory of differentiated and mutually exclusive verbal and spatial STM resource pools, it was hypothesized that shorter RLs should be obtained for class comparisons that included digits, letters, or words for one class, and colors or shapes for the other class. The underlying basis for this hypothesis is that these class combinations use the different STM resource pools in a comparatively more efficient manner which translates into shorter RLs. Results of the experiment do not appear to support this hypothesis. Several interpretations of this finding are offered.

First, it may be the case that colors and shapes lend themselves readily to verbal coding (i.e., attaching a verbal label to a color or a shape) in addition to spatial coding. Verbal coding, however, may represent the dominant or primary coding means and was applied on the stimulus class items used in this experiment. Note that studies citing the use of spatial codes employed stimuli that do not readily lend themselves to a verbal code (e.g., Baddeley and Hitch, 1974). It is, therefore, plausible that the use of other stimulus classes (e.g., random forms) might more clearly bring out the stimulus class processing code effect. During debriefing, participants did comment that they employed some degree of spatial coding for colors and shapes especially when these were presented in the latter part of the memory list. They indicated that this was not used pervasively over verbal coding, however.

As a second interpretation, spatial coding may have been used, but its effect on RL may have been overshadowed by the large processing rate differences among the stimulus classes. For example, shapes elicited the largest RLs and SWAT ratings in all of its comparisons (see Table F16). It is plausible that despite the availability and use of spatial resources, shapes may have been inherently more difficult and time consuming to process than the other stimulus classes. Digits, on the other hand, may have been easier to process even though comparatively fewer verbal resources were available.

The third interpretation considers that spatial resources were employed only when verbal resources were exhausted. This contention is similar to the first interpretation presented above, in that verbal coding may represent the primary means of processing stimulus class information while spatial coding is a secondary means. Following this interpretation, MSET sizes of four or six may not have compelled the use of spatial resources because verbal resources were still available. However, even the MSET size of eight did not manifest any differential processing code effect. Perhaps the use of larger MSET sizes is warranted.

The above interpretations assume means of flexible resource allocation with respect to the processing of stimulus class items. Results from this experiment support the first and second interpretations (i.e., results suggest the use of dual codes for colors and shapes, and large processing rate differences overshadowing the differential processing code effect). Given the goals and objectives of this research program, however, either interpretation does not affect the summary conclusions.

Experiment 2 revealed that verbal/spatial codes are not determining factors in the processing of digits, letters, words, colors, and geometrical shapes when verbal coding schemes are prevalent and when the number of display items is small. Item recognition error increased in a similar fashion with increasing recognition latency. Subjective assessment of processing difficulty dissociated with processing time in a number of class comparisons.

GENERAL DISCUSSION

Two experiments were conducted to investigate memory loading as a function of stimulus class. Experiment 1 employed a tracking-primary/Sternberg-secondary task paradigm, while Experiment 2 employed a modified version of the Sternberg paradigm (Sternberg, 1966). Measures of recognition latency and error and the SWAT technique were used to assess loading for different numbers of items retained and processed in STM.

This research has provided several lessons in developing a better understanding of information display design and cognitive workload (CWL) research. In particular, the following benefits accrued:

1. The use of words, colors, and geometrical shapes provided an insightful and novel extension to employing Sternberg's procedure in CWL research. Previous research of this nature tested digits and letters.
2. In addition to being a sensitive measure of MSET size loading, SWAT also was sensitive to differences in memory loading across stimulus class. Points 1 and 2 further demonstrated the vital link between STM demand and perceived workload.
3. Results of Experiment 1 support the concept of a *differential processing efficiency* across stimulus class. The findings of perceived workload and recognition latency differences in response to items from different stimulus classes opens up the possibility of its use as an information portrayal optimization technique. Substantial recognition time savings in the magnitude of 100 ms to 150 ms can be accrued by presenting a piece of information in the form of a word or a digit as opposed to a color or a shape.

At least in the laboratory environment, SWAT was a more sensitive measurement technique than recognition latency or error rate (compare Tables D11 and D17; see also

Table F16). Error rates did not reveal much in the way of differential processing. This result was expected, partly because STM capacity was not exceeded in most stimulus conditions.

As mentioned previously, the obtained ordering of stimulus class processing rates is crucial to any recommendations that can be derived from this research. To the extent that the ordering is stable, generalizations can be made. The findings of Experiment 1 support the theoretical position of a stable ordering. Except for the discrepancy between the Words and Colors stimulus classes, the results agree with the findings of Brown and Kirsner (1980) and Cavanagh (1972). This discrepancy is likely due to the inclusion of the primary spatial tracking task. Verification of this hypothesis is recommended as future research.

The stimulus class processing rate differences are large enough to warrant testing in context-specific applications. It is beneficial that the procedures and methodologies used in this research are similar to the human-display interaction of many applied environments. For example, aviation operations involve aircraft handling analogous to the tracking task used. A great deal of encoding, storing, and processing of information coded in different stimulus classes are also involved. These processes, similarly, draw upon STM resources. Commercial aviation operations, in particular, depend upon visual information recognition of systems and flight status as words and radar envelopes as colors.

It follows logically that the results of this research should be transferrable to operational environments that employ similar task procedures. This statement should not be taken to mean that testing in applied contexts are not required. As was discussed in the literature review, a number of factors influence stimulus class processing rate (e.g., the level of familiarity or practice, the presentation mode, emphasis on speed over accuracy). Because the level of these factors used in this research may be different from those of applied contexts, verification and testing before implementation is necessary. However, as previously mentioned, it is this author's contention that their moderating effects on processing rate are largely confined to the laboratory environment.

The findings of this research are relevant, especially, in aviation operations because of the extreme STM and cognitive demands typically involved. For example, it was mentioned that savings in the magnitude of 100 ms to 150 ms can be accrued from presenting a piece of information in the form of a word or a color as opposed to a shape. These differences may not seem significant in ordinary human-display interaction, but during the crucial seconds of aircraft collision/avoidance, the milliseconds saved may determine whether or not a disaster can be avoided. Other situations where stimulus class information is of interest to crew station designers include flight data confirmation, enroute data base/flight updates, flight rerouting, etc. The proper formatting of stimulus class information can facilitate crew members' interpretational and decisional STM processes under these time- and effort-stressed situations. Stimulus class optimization is, thus, highly desirable in these contexts.

Other contexts where this research's findings may be of value include air traffic and other process control operations. This research should prove useful also to the more general area of information display design.

Recommendations for Future Research

Several directions for future research are proposed. First, the literature on multi-dimensional stimulus coding suggests that it may be fruitful to investigate stimulus class optimization for information coded in two or more ways. For example, Kopola (1979) found that pilots were better able to encode target information when the information was coded using color and shape than when it was coded using either stimulus class alone.

Two issues are relevant when considering the use of multi-dimensional stimulus coding. One is the orthogonal variation of information quantity where more information can be transmitted when two or more stimulus classes are used to code the same piece of information. The other is redundancy gain whereby coding via two or more stimulus classes helps ensure information transmission (Wickens, 1984b).

As proposed in the experiment discussions, the use of larger MSET sizes in bi-stimulus class presentation may be a more appropriate condition in which to test the hypothesized differential processing code effect. Finally, the testing of other stimulus classes (e.g., auditory versions of the visual stimuli) may shed further insights for optimizing stimulus class in information display design.

REFERENCES

Aasman, J., Mulder, G., and Mulder, L. J. M. (1987). Operator effort and the measurement of heart-rate variability. *Human Factors, 29,* 161-170.

Baddeley, A. D. (1982). *Your memory: A user's guide.* New York: Basic Books.

Baddeley, A. D. (1984). *Working memory.* Oxford: Clarendon Press.

Baddeley, A. D., Grant, W., Wright, E., and Thomson, N. (1975). Imagery and visual working memory. In P. M. A. Rabbitt and S. Dornic (Eds.), *Attention and performance V* (pp. 205-217). New York: Academic Press.

Baddeley, A. D., and Hitch, G. (1974). Working memory. In G. H. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 8, pp. 47-89). New York: Academic Press.

Baddeley, A. D., and Lieberman, K. (1980). Spatial working memory. In R. S. Nickerson (Ed.), *Attention and performance VIII* (pp. 521-539). Hillsdale, NJ: Lawrence Erlbaum Associates.

Battig, W. F., and Montague, W. E. (1969). Category norms for verbal items in 56 categories. *Journal of Experimental Psychology Monograph Supplement, 80.*

Beatty, J. (1982). Task-evoked pupillary responses, processing load, and the structure of processing resources. *Psychological Bulletin, 91,* 276-292.

Brener, R. (1940). An experimental investigation of memory span. *Journal of Experimental Psychology, 26,* 467-482.

Brown, H. L., and Kirsner, K. (1980). A within-subjects analysis of the relationship between memory span and processing rate in short-term memory. *Cognitive Psychology, 12,* 177-187.

Burrows, D., and Okada, R. (1971). Serial position effects in high-speed memory search. *Perception and Psychophysics, 10,* 305-308.

Cavanagh, J. P. (1972). Relationship between the immediate memory span and the memory search rate. *Psychological Review, 79,* 525-530.

Dvorine, I. (1953). *Dvorine pseudo-isochromatic plates.* The Psychological Corporation.

Egeth, H. E. (1969). Parallel versus serial processes in multidimensional stimulus discrimination. In R. N. Haber (Ed.), *Information-processing approaches to visual perception* (pp. 230-240). New York: Holt, Reinhart, and Winston.

Egeth, H. E., and Smith, E. E. (1967). On the nature of errors in a choice reaction task. *Psychonomic Science, 8,* 345-346.

Eggemeier, F. T. (1988). Properties of workload assessment techniques. In P. A. Hancock and N. Meshkati (Eds.), *Human mental workload* (pp. 41-62). Amsterdam: Elsevier Science.

Eggemeier, F. T., Crabtree, M. S., and LaPointe, P. A. (1983). The effects of delayed report on subjective ratings of mental workload. In *Proceedings of the Human Factors Society 27th Annual Meeting* (pp. 139-143). Santa Monica, CA: Human Factors Society.

Eggemeier, F. T., Crabtree, M. S., Zingg, J. J., Reid, G. B., and Shingledecker, C. A. (1982). Subjective workload assessment in a memory update task. In *Proceedings of the Human Factors Society 26th Annual Meeting* (pp. 643-647). Santa Monica, CA: Human Factors Society.

Eggemeier, F. T., and Stadler, M. A. (1984). Subjective workload assessment in a spatial memory task. In *Proceedings of the Human Factors Society 28th Annual Meeting* (pp. 215-219). Santa Monica, CA: Human Factors Society.

Eggleston, R. G., and Quinn, T. J. (1984). A preliminary evaluation of a projective workload assessment procedure. In *Proceedings of the Human Factors Society 28th Annual Meeting* (pp. 695-699). Santa Monica, CA: Human Factors Society.

Frick, R. W. (1984). Using both an auditory and a visual short-term store to increase digit span. *Memory and Cognition, 12*, 507-514.

Gopher, D., Chillag, N., and Arzi, N. (1985). The psychophysics of workload: A second look at the relationship between subjective measures and performance. In *Proceedings of the Human Factors Society 29th Annual Meeting* (pp. 640-644). Santa Monica, CA: Human Factors Society.

Hart, S. G., and Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In P. A. Hancock and N. Meshkati (Eds.), *Human mental workload* (pp. 139-183). Amsterdam: Elsevier Science.

Hart, S. G., and Wickens, C. D. (1990). Workload assessment and prediction. In H. R. Booher (Ed.), *An approach to systems integration* (pp. 257-296). New York: Holt, Reinhart, and Winston.

Hemmingway, J. C. (1984). *An experimental evaluation of the Sternberg task as a workload metric for helicopter flight handling qualities research* (NASA Tech. Memo 85884). Washington, DC: NASA.

Hitch, G. (1984). Working memory. *Psychological Medicine, 14*, 265-271.

Jex, H. R. (1988). Measuring mental workload: Problems, progress, and promises. In P. A. Hancock and N. Meshkati (Eds.), *Human mental workload* (pp. 5-39). Amsterdam: Elsevier Science.

Jex, H. R., and Clement, W. F. (1979). Defining and measuring perceptual-motor workload in a manual control task. In N. Moray (Ed.), *Mental workload: Its theory and measurement* (pp. 125-179). New York: Plenum Press.

Jex, H. R., McDonnell, J. D., and Phatak, A. V. (1966). A critical tracking task for manual control research. *IEEE Transactions (HFE-7), 4*, 138-145.

Johnson, R. M. (1973). Pairwise nonmetric multidimensional scaling. *Psychometrika, 38,* 11-18.

Joula, J. F., and Atkinson, R. C. (1971). Memory scanning for words versus categories. *Journal of Verbal Learning and Verbal Behavior, 10,* 522-527.

Klatzky, R. L. (1980). *Human memory: Structures and processes.* San Francisco, CA: Freeman and Company.

Klatzky, R. L., Joula, J. F., and Atkinson, R. C. (1971). Test stimulus representation and experimental context effects in memory scanning. *Journal of Experimental Psychology, 87,* 281-288.

Kopola, C. (1979). The use of color-coded symbols in a highly dense situation display. In *Proceedings of the Human Factors Society 23rd Annual Meeting.* Santa Monica, CA: Human Factors Society.

Kruskal, J. B. (1965). Analysis of factorial experiments by estimating monotone transformations of the data. *Journal of the Royal Statistical Society, Series B, 27,* 251-263.

Kuperman, G. G., and Wilson, D. L. (1986). *A human-centered technology for advanced bomber crew stations.* (Tech. Report No. AAMRL-TR--86-052). Dayton, OH: Armstrong Aerospace Medical Research Laboratory, Wright Patterson AFB.

Kuperman, G. G., and Wilson, D. L. (1988). The design of a tactical situation display. In *Proceedings of the Human Factors Society 32nd Annual Meeting* (pp. 111-115). Santa Monica, CA: Human Factors Society.

Lachman, R., Lachman, J. L., and Butterfield, E. C. (1979). *Cognitive psychology and information processing: An introduction.* New York: Lawrence Erlbaum Associates.

Lysaght, R. J., Hill, S. G., Dick, A. O., Plamondon, B. D., Linton, P. M., Wierwille, W. W., Zaklad, A. L., Bittner, A. C. Jr., and Wherry, R. J. (1989). *Operator* *workload: Comprehensive review and evaluation of operator workload methodologies.* (Tech. Report No. 851). Alexandria, VA: U. S. Army Research Institute for the Behavioral and Social Sciences.

Meister, D. (1985). *Behavioral analysis and measurement methods.* New York: John Wiley & Sons.

Meshkati, N. (1988). Towards development of a cohesive model of workload. In P. A. Hancock and N. Meshkati (Eds.), *Human mental workload* (pp. 305-314). Amsterdam: Elsevier Science.

Meshkati, N., and Loewenthal, A. (1988). An eclectic and critical review of four primary mental workload assessment methods: A guide for developing a comprehensive model. In P. A. Hancock and N. Meshkati (Eds.), *Human mental workload* (pp. 251-266). Amsterdam: Elsevier Science.

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review, 63,* 81-97.

Neter, J., and Wasserman, W. (1974). *Applied linear statistical models.* Homewood, IL: Richard Irwin.

O'Donnell, R. D., and Eggemeier, F. T. (1986). Workload assessment methodology. In K. R. Boff, L. Kaufman, and J. P. Thomas (Eds.), *Handbook of perception and human performance* (Vol. II, pp. 42.1-42.52). New York: John Wiley & Sons.

Poulton, E. C. (1974). *Tracking skill and manual control.* New York: Academic Press.

Reid, G. B. (1985). Current status on the development of Subjective Workload Assessment Technique. In *Proceedings of the Human Factors Society 29th Annual Meeting* (pp. 220-223). Santa Monica, CA: Human Factors Society.

Reid, G. B., and Nygren, T. E. (1988). The subjective workload assessment technique: A scaling procedure for measuring mental workload. In P. A. Hancock and N. Meshkati (Eds.), *Human mental workload* (pp. 185-218). Amsterdam: Elsevier Science.

Reid, G. B., Potter, S. S., and Bressler, J. R. (1989). *Subjective workload assessment technique (SWAT): A user's guide (U).* (Tech. Report No. AAMRL-TR--89-023). Dayton, OH: Armstrong Aerospace Medical Research Laboratory, Wright-Patterson AFB.

Reid, G. B., Shingledecker, C. A., and Eggemeier, F. T. (1981). Application of conjoint measurement to workload scale development. In *Proceedings of the Human Factors Society 25th Annual Meeting* (pp. 522-526). Santa Monica, CA: Human Factors Society.

Reinhart, W. F., Glynn, C. D., Dye, C., Takahama, M. S., and Snyder, H. L. (1988). *The role of short-term memory in operator workload* (Tech. Report No. AAMRL-TR--88-024). Dayton, OH: Armstrong Aerospace Medical Research Laboratory, Wright-Patterson AFB.

Rouse, W. B. (1979). Approaches to mental workload. In N. Moray (Ed.), *Mental workload: Its theory and measurement* (pp. 255-262). New York: Plenum Press.

Sanders, M. S., and McCormick, E. J. (1987). *Human factors in engineering and design.* New York: McGraw-Hill.

Schneider, W., and Shiffrin, R. M. (1977). Controlled and automatic human information processing: I. Detection, search, and attention. *Psychological Review, 84,* 1-66.

Shiffrin, R. M., and Schneider, W. (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending, and a general theory. *Psychological Review, 84,* 127-190.

Smith, E. E. (1967). Effects of familiarity on stimulus recognition and categorization. *Journal of Experimental Psychology, 74*, 324-332.

Sternberg, S. (1966). High-speed scanning in human memory. *Science*, 153, 652-654.

Sternberg, S. (1967). Retrieval of contextual information from memory. *Psychonomic Science, 8*, 55-56.

Sternberg, S. (1969). Memory scanning: Mental processes revealed by reaction-time experiments. *American Scientist, 57*, 421-427.

Sternberg, S. (1975). Memory scanning: New findings and current controversies. *Quarterly Journal of Experimental Psychology, 27*, 1-32.

Tsang, P. S., and Wickens, C. D. (1984). The effects of task structure on time-sharing efficiency and resource allocation optimality. In *Proceedings of the Twentieth Annual Conference on Manual Control* (Vol. 2, pp. 305-317). Washington, DC: NASA.

Vidulich, M. A., and Tsang, P. S. (1985). Techniques of subjective workload assessment: A comparison of SWAT and NASA-Bipolar methods. *Ergonomics, 29*, 1385-1398.

Wickens, C. D. (1980). The structure of attentional resources. In R. S. Nickerson (Ed.), *Attention and performance VIII* (pp. 239-257). Hillsdale, NJ: Lawrence Erlbaum Associates.

Wickens, C. D. (1984a). Processing resources in attention. In R. Parasuraman and R. D. Davies (Eds.), *Varieties of attention* (pp. 63-102). Hillsdale, NJ: Lawrence Erlbaum Associates.

Wickens, C. D. (1984b). *Engineering psychology and human performance*. Columbus, OH: Charles E. Merrill Publishing.

Wickens, C. D. (1989). Attention and skilled performance. In D. H. Holding (Ed.), *Human skills* (pp. 71-105). New York: John Wiley & Sons.

Wickens, C. D., Hyman, F. C., Dellinger, J. A., Taylor, H. L., and Meador, M. (1986). The Sternberg memory search task as an index of pilot workload. *Ergonomics, 29*, 1371-1383.

Wiener, E. L., and Nagel, D. C. (Eds.). (1988). *Human factors in aviation.* New York: Academic Press.

Wierwille, W. W. (1979). Physiological measures of aircrew mental workload. *Human Factors, 21*, 575-593.

Williges, R. C., and Wierwille, W. W. (1979). Behavioral measures of aircrew mental workload. *Human Factors, 21*, 549-574.

Wilson, G. F., and O'Donnell, R. D. (1988). Measurement of operator workload in the neuropsychological test battery. In P. A. Hancock and N. Meshkati (Eds.), *Human mental workload* (pp. 63-100). Amsterdam: Elsevier Science.

Yeh, Y. Y., and Wickens, C. D. (1988). Dissociation of performance and subjective measures of workload. *Human Factors, 30*, 111-120.

APPENDIX A

PARTICIPANT'S INFORMED CONSENT

PARTICIPANT'S INFORMED CONSENT

You are being asked to participate in a research project. The purpose of this experiment is to examine performance given visual display tasks of different complexities. The task will require you to complete numerous trials. In each trial you will view a computer display and recall displayed information.

This research is being conducted by the Human Factors Laboratory of the Department of Industrial Engineering and Operations Research (IEOR). IEOR research team members on this project include:

Dr. Robert J. Beaton, Faculty Member and Principle Investigator (231-5936)
Kay C. Tan, Doctoral Student and Research Assistant (231-5499)
Michael J. Kahn, Doctoral Student and Research Assistant (231-5499)

This study consists of two parts: screening tests and experiment sessions.

Screening Tests

First, you will complete a screening procedure involving a contrast sensitivity and a color test. This test will last about 10 minutes. You will *not* be paid for completing the screening procedure.

If you are not invited to participate in the second portion of the experiment, you should *not* consider this as an inadequacy. We are seeking people with specific visual abilities. Also, as you are aware, people employ different perceptual styles to solve problems. We are seeking people who use a specific style. In fact, it is expected that many of the volunteers who apply will *not* be appropriate for this specific visual experiment.

Experiment Sessions

The second portion of the study is expected to last 5 days (starting today). The first session will last approximately 2 hours; subsequent sessions will be about 1 hour each. If you decide to participate, you will be paid $25 upon completing all 5 sessions. *If you do not complete all 5 sessions, you will not be paid.* This is because your data will be of no use to us.

During these sessions you will complete numerous trials. In each trial you will view a computer display and recall displayed information. There are no known risks associated with this research. The only known discomfort to which you will be exposed is possible fatigue resulting from the length of the experiment. However, you will be permitted to take rest breaks when needed.

As a Subject, You Have Certain Rights:

1. It is your right to withdraw from the study at any time for any reason. However, as stated earlier, you will *not* be paid if you do not complete all 5 sessions.

2. One of the research team members will answer any questions you may have concerning this research. You should not sign this consent form until you are satisfied that you understand all the terms involved.

3. You have the right to see your data and withdraw them if you so desire. If you decide to withdraw your data, please inform the experimenter immediately. Otherwise, identification of your data will not be possible since it is separated from the participant in order to ensure anonymity.

4. If you wish to receive a summary of the results of this research, please include your address (where you expect to be living three months from now) with your signature below. Please do so only if you are truly interested in seeing the results. If you desire more detailed information after receiving the synopsis, please contact the Human Factors Laboratory and a full report will be made available to you.

5. Should any further questions or problems arise, you may contact any of the research team members. If you have any concerns about the way the experiment is being conducted or the way you are being treated, you may contact Dr. E. R. Stout, Chairperson of the Institutional Review Board (231-5281).

Your participation is greatly appreciated and we hope that you will find the study a pleasant and interesting experience. Your signature below indicates that you have read this document in its entirety, that your questions have been answered, and that you consent to participate in the study described. Furthermore, your signature below indicates that you will not discuss this study with anyone until June, 1990 when the study is to be completed.

____

Please print your name and address if you would like a synopsis of the results

APPENDIX B

SUBJECTIVE WORKLOAD ASSESSMENT TECHNIQUE (SWAT) CARD-SORT INSTRUCTIONS

SWAT CARD-SORT INSTRUCTIONS

PERFORMANCE MEASURES

We are interested in your performance on a number of tasks. Performance will be measured using task completion times and error rates. We also want your assessment of the amount of mental workload associated with each task.

MENTAL WORKLOAD

Mental workload is a measure of mental effort. It is a measure of how hard you have to concentrate or work to complete a mental problem.

* For example, the amount of mental work required to solve a complex arithmetic problem is greater than that required to solve a simple one.

* We want to know how much and what kind of mental work you have to do to complete our tasks.

* One method of measuring mental workload is called SWAT (Subjective Workload Assessment Technique).

SWAT

SWAT describes mental workload in terms of three dimensions:

Time Load

Amount of time pressure. Includes proportion of time you are busy, the degree to which tasks interfere with one another, and/or if there is not enough time to finish task(s).

* With a low Time Load, there is enough time to complete all mental work, plus some time to spare.

* With a high Time Load, you are very busy and task interruptions occur very frequently.

Time Load may be rated according to these *three* levels:

1. Often have spare time. Interruptions or overlap among activities occur infrequently or not at all.

2. Occasionally have spare time. Interruptions or overlap among activities occur frequently.

3. Almost never have spare time. Interruptions or overlap among activities are very frequent, or occur all the time.

Mental Effort Load

Amount of concentration required. Includes memorizing, performing calculations, or making difficult decisions.

- Regardless of the number of tasks that are performed at the same time.
- Regardless of any time limitations.

Mental Effort Load may be rated according to these *three* levels:

1. Very little conscious mental effort or concentration required. Activity is almost automatic, requiring little or no attention.
2. Moderate conscious mental effort or concentration required. Complexity of activity is moderately high due to uncertainty, unpredictability, or unfamiliarity. Considerable attention required.
3. Extensive mental effort and concentration are necessary. Very complex activity requiring total attention.

Psychological Stress Load

How performance on a task is affected by factors such as anxiety, frustration, or confusion.

- At low levels, you feel relatively relaxed.
- As anxiety, frustration, or confusion increases, greater levels of concentration and determination are needed to perform the task.

Psychological Stress Load may be rated according to these *three* levels:

1. Little confusion, risk, frustration, or anxiety exists and can be easily accommodated.
2. Moderate stress due to confusion, frustration, or anxiety noticeably adds to workload. Significant compensation is required to maintain adequate performance.
3. High to very intense stress due to confusion, frustration, or anxiety. High to extreme determination and self-control required.

CARD SORT PROCEDURE

Here is a stack of cards. Each card represents one level of mental workload combined from each of the three SWAT dimensions. For example, level 1 on Time Load, level 2 on Mental Effort Load, and level 2 on Psychological Stress Load combine to make one level of mental workload. Another level of mental workload would be represented by the combination: level 3 on Time Load, level 2 on Mental Effort Load, and level 3 on Psychological Stress Load. Altogether, there are 27 cards representing the 27 combinations (i.e., 3 dimensions x 3 levels each) of the SWAT workload scale.

As part of the experiment, we need you to sort these 27 cards in the order in which *you* perceive as more or less mental workload. Sort in an increasing order from the lowest workload level to the highest workload level. Below is an example of a card sort and the resulting workload scale that will be obtained:

TIME	EFFORT	STRESS		WORKLOAD SCALE
1	1	1	------------>	0.0
.	.	.		.
.	.	.		.
.	.	.		.
3	3	3	------------>	100.0

Points to Remember when doing the Card Sort

1. When doing the card sort, use the descriptions printed on the cards. Use your general view of workload and consider the order of importance of each of the Time, Mental Effort, and Psychological Stress dimensions. Do not sort the cards based only on one task -- use your general view of mental workload.

2. During the experiment, you will be rating the mental workload involved in a series of trials. You will base you ratings on these three dimensions. For example, you may rate the mental workload involved in a particular trial as 1-2-2. This means 1 for Time, 2 for Mental Effort, and 2 for Psychological Stress.

3. We are not asking what you prefer. Some people prefer to be "busy" rather than "idle" in the Time dimension. Again, this is not what we are interested in. We want to know how the three dimensions and their three levels represent mental workload as *you* see it. You may prefer a 2-2-2 situation to a 1-1-1 situation which indicates that you prefer not to have too much spare time or that you work better under stress. Regardless, realize that 1-1-1 is *less* workload than 2-2-2.

4. Please feel free to ask any questions as you are sorting the cards.

APPENDIX C

INSTRUCTIONS FOR EXPERIMENT 1

INSTRUCTIONS FOR EXPERIMENT 1

Thank you for agreeing to participate in this experiment. During the course of this experiment, you will be performing *two* tasks at the *same* time. One is a tracking task and the other is a task involving the use of your memory. Speed, and more importantly accuracy, are of the essence in this experiment. Let us go through exactly what your participation in this experiment is all about. First, the tracking task.

Tracking Task

For this task, you see the joystick on the table, the computer screen with a horizontal line and an inverted triangle on top of this line. During the course of the experiment, the triangle will be moving about constantly; your job is to keep it as close as possible to the center mark at all times. You will do this by moving the joystick with your *preferred* hand. In a few moments, you will get to practice this task.

Memory Task

The other task requires you to remember several items that will be presented on the computer screen -- let's call these items: stimuli. A total of from 2 to 5 stimuli will appear one at a time. Two seconds after all the stimuli have been shown, another item will appear -- we call this item the probe. Your job is to indicate whether the probe was or was not one of the stimuli previously shown. If it is, press the YES button you see on the box lying on the table. If it is not, press the NO button. Since you will be tracking with one hand, use the other hand to press the YES/NO keys. The stimuli may be any of the following types: digits, letters, colors, words, or geometrical shapes. You shall get to practice this task in a few moments.

Trial Sequence

Again, this experiment requires performing both the tracking task and the memory task at the same time. This experiment consists of a series of trials. The sequence of each trial is as follows:

1. You press the YES key on the box to begin a trial.
2. Immediately, begin tracking and try to keep the triangle at the center of the line.
3. About 10 seconds later, a cross will appear at the location where the stimuli will be presented. The stimuli will appear one at a time. After they have been presented, there will be a two second pause, and then the probe item will appear.
4. You respond whether the probe was or was not part of the group of stimuli just presented by pressing either the YES or NO key.
5. After giving your response, you will rate the amount of mental effort involved in completing the trial. Remember, we went through the SWAT scale earlier on. The SWAT dimensions will appear on the computer screen. You make your selections by moving the highlight to the appropriate numbers and then press the YES key.

The above steps constitute one trial. Whenever you are ready, press the YES key to begin the next trial. There will be 80 of the trials just described.

Practice

For this session, you get to go through all 80 trials, but only as practice. We will not be collecting data yet. Actual data collection will begin on the next session.

After Practice

These practice trials are exactly what you will be doing throughout the experiment. The only variation is that the stimuli may appear in a different sequence. *Remember, speed is important, but accuracy in giving the correct response is more important.*

Before we conclude for the day, the members of this research team and our sponsors appreciate your invaluable participation in this experiment. Do you have any questions?

DAILY REVIEW INSTRUCTIONS FOR EXPERIMENT 1

Before beginning today's session, you will have a few practice trials to refresh your memory of the tasks to be performed.

1. *Tracking Task.* Your primary responsibility is to keep the inverted triangle as close as possible at the center of the horizontal line.

2. *Memory Task.* Remember that after the list of stimuli has been presented (2, 3, 4, or 5 stimuli), another item -- the probe -- will appear. You are to determine if the probe was or was not part of the group of stimuli just presented. If it is, press the "YES" key on the box; if it is not, press the "NO" key.

   Again, speed is important but it is more important that you do not make a mistake.

3. *SWAT Ratings.* After you have made your response, the SWAT dimensions will appear. Based on the three dimensions of Time Load, Mental Effort Load, and Psychological Stress Load, rate the amount of mental workload that you experienced in completing the trial.

BEFORE YOU BEGIN YOUR PRACTICE TRIALS, DO YOU HAVE ANY QUESTIONS ABOUT THESE INSTRUCTIONS OR THE TASKS?

APPENDIX D

STATISTICAL ANALYSES RESULTS FOR EXPERIMENT 1

TABLE D1

Experiment 1: Summary of ANOVA Results for Secondary Task Recognition Latency; Data from Blocks 3 through 10

Source	df	MS	*F*	*p*
Subject	9	8.6067		
Block	7	2.9336	7.43	.0001
Block*Subj	63	0.3951		
Stimulus Class	4	6.0962	23.32	.0001
Stimulus Class*Subj	36	0.2615		
MSET Size	3	3.5160	12.28	.0001
MSET Size*Subj	27	0.2862		
Probe Type	1	0.6045	2.97	.1188
Probe Type*Subj	9	0.2034		
Block*Class	28	0.1925	1.21	.2194
Block*Class*Subj	252	0.1588		
Block*Size	21	0.1590	1.06	.3995
Block*Size*Subj	189	0.1506		
Block*Probe	7	0.1373	1.28	.2754
Block*Probe*Subj	63	0.1074		
Class*Size	12	0.2971	2.04	.0268
Class*Size*Subj	108	0.1453		
Class*Probe	4	0.2076	1.34	.2734
Class*Probe*Subj	36	0.1547		
Size*Probe	3	0.2354	1.78	.1754
Size*Probe*Subj	27	0.1325		
Block*Class*Size	84	0.1765	1.07	.3299
Block*Class*Size*Subj	756	0.1655		
Block*Class*Probe	28	0.1523	1.00	.4752
Block*Class*Probe*Subj	252	0.1528		
Block*Size*Probe	21	0.2407	1.59	.0539
Block*Size*Probe*Subj	189	0.1509		

TABLE D1 (continued)

Source	df	MS	F	p
Class*Size*Probe	12	0.1888	1.11	.3568
Class*Size*Probe*Subj	108	0.1695		
Block*Class*Size*Probe	84	0.1599	1.14	.1924
Block*Class*Size*Probe*Subj	756	0.1401		
Total	3119			

TABLE D2

Experiment 1: Summary of ANOVA Results for Secondary Task Recognition Error; Data from Blocks 3 through 10

Source	df	MS	*F*	*p*
Subject	9	0.1283		
Block	7	0.0279	0.81	.5788
Block*Subj	63	0.0342		
Stimulus Class	4	0.1458	4.98	.0027
Stimulus Class*Subj	36	0.0293		
MSET Size	3	0.0108	0.37	.7759
MSET Size*Subj	27	0.0294		
Probe Type	1	0.0613	1.52	.2495
Probe Type*Subj	9	0.0404		
Block*Class	28	0.0233	0.73	.8419
Block*Class*Subj	252	0.0320		
Block*Size	21	0.0356	1.07	.3817
Block*Size*Subj	189	0.0332		
Block*Probe	7	0.0513	1.56	.1631
Block*Probe*Subj	63	0.0328		
Class*Size	12	0.0324	1.13	.3430
Class*Size*Subj	108	0.0287		
Class*Probe	4	0.0214	0.69	.6027
Class*Probe*Subj	36	0.0310		
Size*Probe	3	0.0321	1.23	.3177
Size*Probe*Subj	27	0.0261		
Block*Class*Size	84	0.0328	1.06	.3441
Block*Class*Size*Subj	756	0.0310		
Block*Class*Probe	28	0.0346	1.17	.2641
Block*Class*Probe*Subj	252	0.0297		
Block*Size*Probe	21	0.0373	1.01	.4502
Block*Size*Probe*Subj	189	0.0369		

TABLE D2 (continued)

Source	df	MS	*F*	*p*
Class*Size*Probe	12	0.0547	1.51	.1319
Class*Size*Probe*Subj	108	0.0363		
Block*Class*Size*Probe	84	0.0260	0.78	.9221
Block*Class*Size*Probe*Subj	756	0.0333		
Total	3119			

TABLE D3

Experiment 1: Summary of ANOVA Results for Secondary Task SWAT Ratings; Data from Blocks 3 through 10

Source	df	MS	*F*	*p*
Subject	9	37196.1552		
Block	7	1320.4073	3.54	.0028
Block*Subj	63	372.6383		
Stimulus Class	4	15285.8102	12.72	.0001
Stimulus Class*Subj	36	1201.8490		
MSET Size	3	190966.9045	52.24	.0001
MSET Size*Subj	27	3655.5677		
Probe Type	1	243.6528	6.49	.0314
Probe Type*Subj	9	37.5710		
Block*Class	28	210.6471	0.98	.4935
Block*Class*Subj	252	214.1953		
Block*Size	21	335.6461	1.35	.1472
Block*Size*Subj	189	248.2122		
Block*Probe	7	230.5621	1.56	.1636
Block*Probe*Subj	63	147.6939		
Class*Size	12	848.0328	3.67	.0001
Class*Size*Subj	108	231.3536		
Class*Probe	4	268.1530	1.40	.2545
Class*Probe*Subj	36	191.9113		
Size*Probe	3	125.9113	0.91	.4487
Size*Probe*Subj	27	138.2733		
Block*Class*Size	84	191.7530	1.04	.3768
Block*Class*Size*Subj	756	183.5389		
Block*Class*Probe	28	161.5614	0.91	.5932
Block*Class*Probe*Subj	252	176.5946		
Block*Size*Probe	21	195.9965	1.33	.1608
Block*Size*Probe*Subj	189	147.4825		

TABLE D3 (continued)

Source	df	MS	*F*	*p*
Class*Size*Probe	12	178.2714	1.01	.4413
Class*Size*Probe*Subj	108	175.7943		
Block*Class*Size*Probe	84	153.1757	0.96	.5721
Block*Class*Size*Probe*Subj	756	158.9259		
Total	3119			

TABLE D4

Experiment 1: Results of Newman-Keuls Tests on Block Effect for Secondary Task Recognition Latency (ms) and SWAT Ratings*

Block	3	4	5	6	7	8	9	10
Recognition Latency	999	985	917	898	834	811	804	770
SWAT Ratings	27.42	26.74	26.99	24.89	23.85	23.38	23.31	23.13

* Means with a common line are not significantly different at $p < .05$.

TABLE D5

Experiment 1: Partial Summary of ANOVA Results for Secondary Task Recognition Latency; Data from Blocks 7 through 10; Only Effects that include Block as one of its Variables are shown

Source	df	MS	*F*	*p*
Subject	9	3.7967		
Block	3	0.2814	1.48	.2413
Block*Subj	27	0.1897		
Block*Stimulus Class	12	0.0873	1.04	.4221
Block*Stimulus Class*Subj	108	0.0843		
Block*MSET Size	9	0.1619	1.86	.0695
Block*MSET Size*Subj	81	0.0869		
Block*Probe Type	3	0.0498	0.55	.6522
Block*Probe Type*Subj	27	0.0905		
Block*Class*Size	36	0.0787	0.85	.7172
Block*Class*Size*Subj	324	0.0926		
Block*Class*Probe	12	0.0961	1.25	.2620
Block*Class*Probe*Subj	108	0.0772		
Block*Size*Probe	9	0.1191	1.32	.2387
Block*Size*Probe*Subj	81	0.0901		
Block*Class*Size*Probe	36	0.0793	0.96	.5442
Block*Class*Size*Probe*Subj	324	0.0829		
Total	1559			

TABLE D6

Experiment 1: Partial Summary of ANOVA Results for Secondary Task SWAT Ratings; Data from Blocks 7 through 10; Only Effects that include Block as one of its Variables are shown

Source	df	MS	*F*	*p*
Subject	9	20431.3194		
Block	3	37.2994	0.11	.9555
Block*Subj	27	349.8512		
Block*Stimulus Class	12	88.3672	0.56	.8699
Block*Stimulus Class*Subj	108	157.9317		
Block*MSET Size	9	244.9021	1.44	.1849
Block*MSET Size*Subj	81	170.0084		
Block*Probe Type	3	81.8028	0.56	.6471
Block*Probe Type*Subj	27	146.5230		
Block*Class*Size	36	142.0796	0.99	.4941
Block*Class*Size*Subj	324	143.9199		
Block*Class*Probe	12	108.5465	0.67	.7752
Block*Class*Probe*Subj	108	161.6422		
Block*Size*Probe	9	99.6163	0.82	.5992
Block*Size*Probe*Subj	81	121.4526		
Block*Class*Size*Probe	36	125.9439	0.98	.5096
Block*Class*Size*Probe*Subj	324	128.8224		
Total	1559			

TABLE D7

Experiment 1: Summary of ANOVA Results for Secondary Task Intrusion Differences across all combinations of the Experimental Variables of Stimulus Class, Memory-Set Size, and Probe Type; Data from Blocks 7 through 10

Source	df	MS	*F*	*p*
Subject	9	5560.2270		
Stimulus Class	4	33.4496	0.08	.9877
Stimulus Class*Subj	36	412.9734		
MSET Size	3	190.4640	0.88	.4615
MSET Size*Subj	27	215.3033		
Probe Type	1	153.9212	0.54	.4806
Probe Type*Subj	9	284.2530		
Class*Size	12	199.8621	1.03	.4242
Class*Size*Subj	108	193.4156		
Class*Probe	4	426.1959	1.68	.1752
Class*Probe*Subj	36	253.1951		
Size*Probe	3	275.6916	1.51	.2344
Size*Probe*Subj	27	182.5835		
Class*Size*Probe	12	278.3742	1.22	.2757
Class*Size*Probe*Subj	108	227.2887		
Total	399			

TABLE D8

Experiment 1: Summary of ANOVA Results for Secondary Task Recognition Latency

Source	df	MS	*F*	*p*
Subject	9	3.7967		
Stimulus Class	4	2.3214	16.64	.0001
Stimulus Class*Subj	36	0.1476		
MSET Size	3	1.8399	8.62	.0004
MSET Size*Subj	27	0.2135		
Probe Type	1	0.7147	4.44	.0004
Probe Type*Subj	9	0.1611		
Class*Size	12	0.2773	4.12	.0001
Class*Size*Subj	108	0.0673		
Class*Probe	4	0.0705	0.75	.6949
Class*Probe*Subj	36	0.0945		
Size*Probe	3	0.1418	3.41	.0315
Size*Probe*Subj	27	0.0415		
Class*Size*Probe	12	0.1254	1.68	.0814
Class*Size*Probe*Subj	108	0.0747		
Total	399			

TABLE D9

Experiment 1: Results of Newman-Keuls Test on Stimulus Class Effect for Secondary Task Recognition Latency (ms)*

Stimulus Class	Digits	Words	Letters	Colors	Shapes
Recognition Latency	780	803	853	903	997

*Means with a common line are not significantly different at $p < .05$.

TABLE D10

Experiment 1: Results of Newman-Keuls Test on Memory-Set Size Effect for Secondary Task Recognition Latency (ms)*

Memory-Set Size	2	3	4	5
Recognition Latency	787	842	887	950

*All means are significantly different from one another at $p < .05$.

TABLE D11

Experiment 1: Results of Simple-Effects F-Tests and Newman-Keuls Tests on Memory-Set Size Effect on each Stimulus Class for Secondary Task Recognition Latency (ms)*

	Memory-Set Size				
Digits	2	3	4	5	
Recognition Latency	749	767	793	813	$p = .1023$
Words	2	3	4	5	
Recognition Latency	758	793	796	862	$p = .0758$
Letters	2	3	4	5	
Recognition Latency	776	844	863	929	$p = .0748$
Colors	2	3	4	5	
Recognition Latency	810	865	930	1011	$p = .0041$
Shapes	2	3	4	5	
Recognition Latency	842	942	1063	1144	$p < .0001$

*Means with a common line are not significantly different at $p < .05$.

TABLE D12

Experiment 1: Test of Regression Line (Slope) Differences for all Stimulus Class pairs with Recognition Latency as the Dependent Variable*

-13.5053	≤	Letters - Digits	≤	+45.0933
-53.5907	≤	Letters - Colors	≤	+9.2707
-29.4745	≤	Letters - Words	≤	+36.2005
-122.5592	≤	Letters - Shapes	≤	+7.1372
-51.9337	≤	Digits - Colors	≤	-23.9743
-29.3247	≤	Digits - Words	≤	+4.4847
-131.9243	≤	Digits - Shapes	≤	-15.0857
+5.1474	≤	Colors - Words	≤	+45.9006
-93.83	≤	Colors - Shapes	≤	+22.7295
-121.3459	≤	Words - Shapes	≤	-0.8041

* The end points represent confidence limits and an $\propto$ value of .05 was used for these tests (see Equation 2). No sign change indicates statistically different slopes.

TABLE D13

Experiment 1: Summary of ANOVA Results for Secondary Task Recognition Error

Source	df	MS	*F*	*p*
Subject	9	0.1140		
Stimulus Class	4	0.0494	2.21	.0868
Stimulus Class*Subj	36	0.0223		
MSET Size	3	0.0040	0.14	.9373
MSET Size*Subj	27	0.0290		
Probe Type	1	0.0506	1.12	.3168
Probe Type*Subj	9	0.0451		
Class*Size	12	0.0248	0.82	.6263
Class*Size*Subj	108	0.0301		
Class*Probe	4	0.0069	0.27	.9678
Class*Probe*Subj	36	0.0256		
Size*Probe	3	0.0523	1.38	.3921
Size*Probe*Subj	27	0.0504		
Class*Size*Probe	12	0.0523	1.84	.0507
Class*Size*Probe*Subj	108	0.0284		
Total	399			

TABLE D14

Experiment 1: Summary of ANOVA Results for Secondary Task SWAT Ratings

Source	df	MS	*F*	*p*
Subject	9	20431.3194		
Stimulus Class	4	8665.7172	13.58	.0001
Stimulus Class*Subj	36	637.9143		
MSET Size	3	104830.7293	41.07	.0001
MSET Size*Subj	27	2552.6789		
Probe Type	1	50.6944	1.27	.2883
Probe Type*Subj	9	39.8039		
Class*Size	12	654.9422	3.19	.0006
Class*Size*Subj	108	205.1910		
Class*Probe	4	233.4940	1.74	.1625
Class*Probe*Subj	36	134.1515		
Size*Probe	3	31.1264	0.25	.8636
Size*Probe*Subj	27	126.6832		
Class*Size*Probe	12	120.2275	0.84	.6049
Class*Size*Probe*Subj	108	142.3783		
Total	399			

TABLE D15

Experiment 1: Results of Newman-Keuls Test on Stimulus Class Effect for Secondary Task SWAT Ratings*

Stimulus Class	Digits	Letters	Colors	Words	Shapes
SWAT Ratings	18.62	21.54	25.69	26.44	31.00

*Means with a common line are not significantly different at $p < .05$.

TABLE D16

Experiment 1: Results of Newman-Keuls Test on Memory-Set Size Effect for Secondary Task SWAT Ratings*

Memory-Set Size	2	3	4	5
SWAT Ratings	6.06	18.00	30.86	43.70

*All means are significantly different from one another at $p < .05$.

TABLE D17

Experiment 1: Results of Simple-Effects F-Tests and Newman-Keuls Tests on Memory-Set Size Effect on each Stimulus Class for Secondary Task SWAT Ratings*

	Memory-Set Size				
Letters	2	3	4	5	
SWAT Ratings	4.66	15.67	26.34	39.56	$p < .0001$
Digits	2	3	4	5	
SWAT Ratings	4.23	11.47	24.17	34.718	$p < .0001$
Colors	2	3	4	5	
SWAT Ratings	5.95	17.85	32.79	46.84	$p < .0001$
Words	2	3	4	5	
SWAT Ratings	8.03	19.89	32.39	45.04	$p < .0001$
Shapes	2	3	4	5	
SWAT Ratings	7.50	25.19	39.03	52.73	$p < .0001$

* All means within each class are significantly different from one another at $p < .05$.

TABLE D18

Experiment 1: Test of Regression Line (Slope) Differences for all Stimulus Class Pairs with SWAT Ratings as the Dependent Variable*

-1.4940	≤	Letters - Digits	≤	+3.2820
-4.2221	≤	Letters - Colors	≤	+0.2161
-2.3890	≤	Letters - Words	≤	+1.2950
-5.3779	≤	Letters - Shapes	≤	-0.7401
-5.3020	≤	Digits - Colors	≤	-0.4919
-3.5033	≤	Digits - Words	≤	+0.6213
-6.4505	≤	Digits - Shapes	≤	-1.4555
-3.3201	≤	Colors - Words	≤	+0.4081
-3.3925	≤	Colors - Shapes	≤	+1.2805
-4.4939	≤	Words - Shapes	≤	-0.5301

* The end points represent confidence limits and an $\propto$ value of .05 was used for these tests (see Equation 2). No sign change indicates statistically different slopes.

TABLE D19

Experiment 1: Pearson's Product Moment correlations among the variables of Memory-Set Size, Recognition Latency, and SWAT Ratings*

	Memory-Set Size	Recognition Latency	SWAT Ratings
Memory-Set Size	. .	0.1618 0.0001	0.6037 0.0001
Recognition Latency	. .	. .	0.2924 0.0001

*All correlation coefficients are significant beyond .0001.

APPENDIX E

INSTRUCTIONS FOR EXPERIMENT 2

INSTRUCTIONS FOR EXPERIMENT 2

Task Description

Thank you for agreeing to participate in this experiment. During the course of this experiment, you will be performing a task involving the use of your memory. *Speed, and more importantly accuracy, are of the essence in this experiment.* Let us go through exactly what your participation in this experiment is all about.

You will be required to memorize several items that will be presented on the computer screen -- let us call these items stimuli. A total of 4, 6, or 8 stimuli will appear one after another. Each stimulus will remain on the screen for 2 seconds. Another 2 seconds after the last stimulus has been shown, another item will appear -- we call this item the probe. Your job is to indicate whether the probe was or was not one of the stimuli previously shown. If it was, press the "YES" button you see on the box lying on the table. If it is not, press the "NO" button. The stimuli may be from any one of the following classes: digits, letters, colors, words, or geometrical shapes. You shall get to practice this task in a few moments.

Trial Sequence

This experiment has been broken-up into segments called trials. Each trial will proceed as follows:

1. You press either the "YES" or "NO" button on the box to begin a trial.

2. *Immediately*, the stimuli will begin appearing on the screen. After the last stimulus has been presented, the screen will be blank for 2 seconds and then the probe will appear. You respond whether the probe was or was not one of the stimuli just presented by pressing either the "YES" or "NO" button.

3. After giving your response, the SWAT dimensions will appear. This is when you rate the amount of mental workload involved in completing the trial.

4. Whenever you are ready, press either the "YES" or "NO" button to begin the next trial.

This experiment consists of 120 of the trials just described.

Practice

For today, you will go through the 120 trials as practice.

After Practice

With the exception of one variation, these practice trials are exactly what you will be encountering throughout the experiment. The stimuli to be presented on each trial will be from two classes. For example, they may be digits and letters, colors and words, or shapes and letters. They will not be digits, letters, and colors, or digits, words, and shapes. Remember, *speed is important, but accuracy is more important.*

Before we begin, the members of this research team and our sponsors appreciate your invaluable participation in this experiment. Do you have any questions?

DAILY REVIEW INSTRUCTIONS FOR EXPERIMENT 2

Before beginning today's session, you will have a few practice trials to refresh your memory of the tasks to be performed.

1. *Memory Task.* You will see either 4, 6, or 8 stimuli on any one trial. They will be from 2 classes; that is, digits and letters, digits and shapes, colors and words, etc. There will be a total of 120 trials with an equal number of trials for each class. After the stimuli have been presented, the probe will appear. You are to determine if the probe was or was not part of the group of stimuli just presented. If it is, press the "YES" key on the box; if it is not, press the "NO" key.

   Again, speed is important but it is more important that you do not make a mistake.

2. *SWAT Ratings.* After you have made your response, the SWAT dimensions will appear. Based on the three dimensions of *Time Load, Mental Effort Load*, and *Psychological Stress Load*, rate the amount of mental workload that you experienced in completing the trial.

BEFORE BEGINNING THE PRACTICE TRIALS, DO YOU HAVE ANY QUESTIONS ABOUT THESE INSTRUCTIONS OR THE TASKS?

APPENDIX F

STATISTICAL ANALYSES RESULTS FOR EXPERIMENT 2

TABLE F1

Experiment 2: Summary of ANOVA Results for Sternberg Task Recognition Latency

Source	df	MS	*F*	*p*
Subject	9	6857893.4122		
Block	3	2458019.8735	5.39	.0049
Block*Subj	27	456424.4496		
ClasComp	9	1508703.5186	8.13	.0001
ClasComp*Subj	81	185634.2903		
ClaSize	2	26355746.2530	47.66	.0001
ClaSize*Subj	18	553010.2790		
PType	1	6045721.5002	5.14	.0496
PType*Subj	9	1176006.9331		
Block*CC	27	129450.5783	0.85	.6763
Block*CC*Subj	243	151450.0826		
Block*CS	6	117897.3167	0.81	.5690
Block*CS*Subj	54	146106.5689		
Block*PT	3	106742.1841	1.07	.3768
Block*PT*Subj	27	99424.3126		
CC*CS	18	207160.0890	1.49	.0988
CC*CS*Subj	162	138878.0288		
CC*PT	9	214663.6683	1.33	.2355
CC*PT*Subj	81	161622.95144		
CS*PT	2	1010687.3658	3.49	.0524
CS*PT*Subj	18	289566.3300		
Block*CC*CS	54	115223.3844	0.86	.7547
Block*CC*CS*Subj	486	134435.9609		
Block*CC*PT	27	145517.0985	0.93	.5702
Block*CC*PT*Subj	243	156651.1946		
Block*CS*PT	6	15311.6031	0.10	.9956
Block*CS*PT*Subj	54	146950.4017		

TABLE F1 (continued)

Source	df	MS	*F*	*p*
CC*CS*PT	18	164788.1769	1.34	.1671
CC*CS*PT*Subj	162	122573.6303		
Block*CC*CS*PT	54	125470.3364	0.92	.6315
Block*CC*CS*PT*Subj	486	135932.8794		
Total	2399			

TABLE F2

Experiment 2: Summary of ANOVA Results for Sternberg Task Recognition Error

Source	df	MS	*F*	*p*
Subject	9	0.1942		
Block	3	0.0624	1.99	.1396
Block*Subj	27	0.0315		
ClasComp	9	0.1099	3.29	.0018
ClasComp*Subj	81	0.0334		
ClaSize	2	0.4215	7.76	.0037
ClaSize*Subj	18	0.0543		
PType	1	0.0469	0.44	.5221
PType*Subj	9	0.1057		
Block*CC	27	0.0274	0.83	.7127
Block*CC*Subj	243	0.0331		
Block*CS	6	0.0645	1.78	.1210
Block*CS*Subj	54	0.0363		
Block*PT	3	0.0485	0.89	.4566
Block*PT*Subj	27	0.0543		
CC*CS	18	0.0222	0.58	.9108
CC*CS*Subj	162	0.0383		
CC*PT	9	0.0168	0.42	.9186
CC*PT*Subj	81	0.0396		
CS*PT	2	0.0694	1.98	.1676
CS*PT*Subj	18	0.0351		
Block*CC*CS	54	0.0501	1.55	.0096
Block*CC*CS*Subj	486	0.0323		
Block*CC*PT	27	0.0333	0.91	.5982
Block*CC*PT*Subj	243	0.0366		
Block*CS*PT	6	0.0335	0.70	.6522
Block*CS*PT*Subj	54	0.0480		

TABLE F2 (continued)

Source	df	MS	*F*	*p*
CC*CS*PT	18	0.0372	1.07	.3891
CC*CS*PT*Subj	162	0.0348		
Block*CC*CS*PT	54	0.0356	1.10	.2936
Block*CC*CS*PT*Subj	486	0.0323		
Total	2399			

TABLE F3

Experiment 2: Summary of ANOVA Results for Sternberg Task SWAT Ratings

Source	df	MS	*F*	*p*
Subject	9	33509.4785		
Block	3	2834.1381	2.24	.1064
Block*Subj	27	1264.9559		
ClasComp	9	10804.6620	25.13	.0001
ClasComp*Subj	81	429.9202		
ClaSize	2	662020.2204	73.90	.0001
ClaSize*Subj	18	8958.4801		
PType	1	2828.5446	3.28	.1035
PType*Subj	9	862.0479		
Block*CC	27	195.6553	1.33	.1337
Block*CC*Subj	243	146.9905		
Block*CS	6	319.3928	0.89	.5085
Block*CS*Subj	54	358.6792		
Block*PT	3	157.9210	1.36	.2767
Block*PType*Subj	27	116.3094		
CC*CS	18	1227.9751	5.73	.0001
CC*CS*Subj	162	214..4501		
CC*PT	9	210.7837	1.06	.4018
CC*PT*Subj	81	199.0219		
CS*PT	2	845.0421	2.91	.0802
CS*PT*Subj	18	290.2032		
Block*CC*CS	54	161.8823	1.25	.1213
Block*CC*CS*Subj	486	129.9245		
Block*CC*PT	27	162.6623	1.04	.4157
Block*CC*PT*Subj	243	156.4384		
Block*CS*PT	6	164.6763	0.96	.4615
Block*CS*PT*Subj	54	171.6758		

TABLE F3 (continued)

Source	df	MS	F	p
CC*CS*PT	18	177.8965	1.03	.4344
CC*CS*PT*Subj	162	173.5033		
Block*CC*CS*PT	54	177.6370	1.06	.3588
Block*CC*CS*PT*Subj	486	166.9640		
Total	2399			

TABLE F4

Experiment 2: Newman-Keuls Test on Block Effect for Recognition Latency*

Block	1	2	3	4
Recognition Latency	918.46	863.70	843.03	810.71

*Means with a common line are not significantly different at $p < .05$.

TABLE F5

Experiment 2: Summary of ANOVA results for the Letters versus Digits Class-Comparison; Dependent Variable is Recognition Latency

Source	df	MS	*F*	*p*
Subject	9	1053741.2120		
ClaSize	2	2186642.7771	13.26	.0003
ClaSize*Subj	18	164852.7516		
PClas	1	409150.4083	10.11	.0112
PClas*Subj	9	40450.3620		
PType	1	924709.6333	7.60	.0222
PType*Subj	9	121617.5315		
ClaSize*PClas	2	226784.2771	1.33	.2902
ClaSize*PClas*Subj	18	170987.5988		
ClaSize*PType	2	140082.9646	1.51	.2467
ClaSize*PType*Subj	18	92508.2586		
PClas*PType	1	387376.0333	2.69	.1354
PClas*PType*Subj	9	144028.3852		
ClaSize*PClas*PType	2	131667.4146	1.72	.2078
ClaSize*PClas*PType*Subj	18	76692.1484		
Total	119			

TABLE F6

Experiment 2: Summary of ANOVA results for the Letters versus Colors Class-Comparison; Dependent Variable is Recognition Latency

Source	df	MS	*F*	*p*
Subject	9	496388.2410		
ClaSize	2	1527837.8396	26.68	.0001
ClaSize*Subj	18	57265.6521		
PClas	1	5514.8521	0.05	.8356
PClas*Subj	9	120851.1299		
PType	1	725018.8021	3.73	.0855
PType*Subj	9	194393.0428		
ClaSize*PClas	2	90459.1896	1.68	.2137
ClaSize*PClas*Subj	18	53740.9604		
ClaSize*PType	2	21551.7896	0.36	.7001
ClaSize*PType*Subj	18	59253.1484		
PClas*PType	1	245933.8021	2.83	.1268
PClas*PType*Subj	9	86917.2373		
ClaSize*PClas*PType	2	14700.8146	0.17	.8492
ClaSize*PClas*PType*Subj	18	89091.5206		
Total	119			

TABLE F7

Experiment 2: Summary of ANOVA results for the Letters versus Words Class-Comparison; Dependent Variable is Recognition Latency

Source	df	MS	*F*	*p*
Subject	9	1262209.9141		
ClaSize	2	4868783.2312	13.05	.0003
ClaSize*Subj	18	373128.9072		
PClas	1	2323387.5521	7.28	.0245
PClas*Subj	9	319205.7604		
PType	1	1917108.8021	2.64	.1383
PType*Subj	9	724903.2697		
ClaSize*PClas	2	1488947.3896	2.29	.1297
ClaSize*PClas*Subj	18	649494.9313		
ClaSize*PType	2	260469.5146	0.53	.5948
ClaSize*PType*Subj	18	487050.6211		
PClas*PType	1	278258.8521	0.61	.4534
PClas*PType*Subj	9	453217.2271		
ClaSize*PClas*PType	2	223548.7896	0.37	.6949
ClaSize*PClas*PType*Subj	18	601786.0674		
Total	119			

TABLE F8

Experiment 2: Summary of ANOVA results for the Letters versus Shapes Class-Comparison; Dependent Variable is Recognition Latency

Source	df	MS	*F*	*p*
Subject	9	907608.0558		
ClaSize	2	3860448.8812	17.12	.0001
ClaSize*Subj	18	225513.9600		
PClas	1	1111206.3021	7.72	.0215
PClas*Subj	9	143928.7188		
PType	1	1116794.6021	5.36	.0458
PType*Subj	9	208343.8243		
ClaSize*PClas	2	33483.5771	0.24	.7891
ClaSize*PClas*Subj	18	139495.9243		
ClaSize*PType	2	377530.7521	3.31	.0597
ClaSize*PType*Subj	18	114097.8076		
PClas*PType	1	343951.6687	21.91	.0012
PClas*PType*Subj	9	15696.1039		
ClaSize*PClas*PType	2	20785.3563	0.28	.7582
ClaSize*PClas*PType*Subj	18	73954.0414		
Total	119			

TABLE F9

Experiment 2: Summary of ANOVA results for the Digits versus Colors Class-Comparison; Dependent Variable is Recognition Latency

Source	df	MS	*F*	*p*
Subject	9	638721.5345		
ClaSize	2	2435778.3583	16.85	.0001
ClaSize*Subj	18	144562.4880		
PClas	1	37789.7521	0.14	.7139
PClas*Subj	9	263894.0345		
PType	1	61630.6687	0.50	.4963
PType*Subj	9	122636.7845		
ClaSize*PClas	2	91052.4333	0.73	.4941
ClaSize*PClas*Subj	18	124138.6602		
ClaSize*PType	2	10897.6000	0.15	.8642
ClaSize*PType*Subj	18	74087.4102		
PClas*PType	1	25623.0188	0.16	.6996
PClas*PType*Subj	9	161404.0419		
ClaSize*PClas*PType	2	243194.0250	1.44	.2640
ClaSize*PClas*PType*Subj	18	169412.9509		
Total	119			

TABLE F10

Experiment 2: Summary of ANOVA results for the Digits versus Words Class-Comparison; Dependent Variable is Recognition Latency

Source	df	MS	*F*	*p*
Subject	9	692210.7120		
ClaSize	2	1099759.4021	15.04	.0001
ClaSize*Subj	18	73108.1127		
PClas	1	41366.5333	0.67	.4347
PClas*Subj	9	61877.5935		
PType	1	448963.3333	4.56	.0616
PType*Subj	9	98560.8750		
ClaSize*PClas	2	14183.8521	0.22	.8070
ClaSize*PClas*Subj	18	65371.7248		
ClaSize*PType	2	23017.6646	0.34	.7182
ClaSize*PType*Subj	18	68278.1576		
PClas*PType	1	94192.0333	1.03	.3374
PClas*PType*Subj	9	91745.4824		
ClaSize*PClas*PType	2	12840.5146	0.16	.8520
ClaSize*PClas*PType*Subj	18	79451.1512		
Total	119			

TABLE F11

Experiment 2: Summary of ANOVA results for the Digits versus Shapes Class-Comparison; Dependent Variable is Recognition Latency

Source	df	MS	*F*	*p*
Subject	9	424868.4910		
ClaSize	2	2829438.0563	21.36	.0001
ClaSize*Subj	18	132447.6326		
PClas	1	1458276.7688	9.74	.0123
PClas*Subj	9	149748.5928		
PType	1	331012.5521	2.24	.1689
PType*Subj	9	147909.4965		
ClaSize*PClas	2	1925.4938	0.02	.9789
ClaSize*PClas*Subj	18	90029.9637		
ClaSize*PType	2	14187.4021	0.34	.7170
ClaSize*PType*Subj	18	41861.4090		
PClas*PType	1	96361.6687	0.76	.4075
PClas*PType*Subj	9	127619.1502		
ClaSize*PClas*PType	2	2015.3313	0.03	.9720
ClaSize*PClas*PType*Subj	18	70859.9725		
Total	119			

TABLE F12

Experiment 2: Summary of ANOVA results for the Colors versus Words Class-Comparison; Dependent Variable is Recognition Latency

Source	df	MS	*F*	*p*
Subject	9	909611.6826		
ClaSize	2	2373362.5187	16.90	.0001
ClaSize*Subj	18	140471.4215		
PClas	1	207875.2521	2.19	.1729
PClas*Subj	9	94842.7475		
PType	1	969931.1021	5.17	.0490
PType*Subj	9	187490.4586		
ClaSize*PClas	2	2441.5396	0.06	.9455
ClaSize*PClas*Subj	18	43422.2294		
ClaSize*PType	2	629323.2021	6.35	.0082
ClaSize*PType*Subj	18	99126.9891		
PClas*PType	1	51067.5021	0.65	.4425
PClas*PType*Subj	9	79123.4234		
ClaSize*PClas*PType	2	160198.7896	4.43	.0273
ClaSize*PClas*PType*Subj	18	36178.8359		
Total	119			

TABLE F13

Experiment 2: Summary of ANOVA results for the Colors versus Shapes Class-Comparison; Dependent Variable is Recognition Latency

Source	df	MS	*F*	*p*
Subject	9	1040058.9732		
ClaSize	2	4057057.8250	22.51	.0001
ClaSize*Subj	18	180200.1190		
PClas	1	1873250.4083	5.26	.0474
PClas*Subj	9	355847.9500		
PType	1	1335630.0000	2.47	.1504
PType*Subj	9	540381.8843		
ClaSize*PClas	2	82576.4583	0.23	.7931
ClaSize*PClas*Subj	18	351695.6597		
ClaSize*PType	2	1014899.1750	4.38	.0281
ClaSize*PType*Subj	18	231591.7468		
PClas*PType	1	1093475.2083	3.38	.0993
PClas*PType*Subj	9	323796.2685		
ClaSize*PClas*PType	2	637397.9083	2.58	.1034
ClaSize*PClas*PType*Subj	18	246915.5449		
Total	119			

TABLE F14

Experiment 2: Summary of ANOVA results for the Words versus Shapes Class-Comparison; Dependent Variable is Recognition Latency

Source	df	MS	*F*	*p*
Subject	9	1103183.2086		
ClaSize	2	2981078.1646	9.57	.0015
ClaSize*Subj	18	311361.4933		
PClas	1	5357934.1021	26.69	.0006
PClas*Subj	9	200711.4956		
PType	1	146895.0187	0.52	.4906
PType*Subj	9	284376.3289		
ClaSize*PClas	2	349032.7646	2.84	.0850
ClaSize*PClas*Subj	18	123050.9359		
ClaSize*PType	2	1820.8938	0.01	.9855
ClaSize*PType*Subj	18	124873.4539		
PClas*PType	1	115103.1021	5.87	.0384
PClas*PType*Subj	9	19604.8475		
ClaSize*PClas*PType	2	153004.5021	0.90	.4245
ClaSize*PClas*PType*Subj	18	170196.9975		
Total	119			

TABLE F15

Experiment 2: Probe Class Recognition Latency and Associated *p*-value for each Class Comparison

Class Comparison	Probe Class	Recognition Latency (in ms)	*p*-value for Probe Class Effect
Letters vs. Digits	Letter	834.79	.0112
	Digit	791.61	
Letters vs. Colors	Letter	827.48	.8356
	Color	827.19	
Letters vs. Words	Letter	911.96	.0245
	Word	776.33	
Letters vs. Shapes	Letter	865.32	.0215
	Shape	957.57	
Digits vs. Colors	Digit	845.88	.7139
	Color	827.49	
Digits vs. Words	Digit	772.83	.4347
	Word	738.09	
Digits vs. Shapes	Digit	794.30	.0123
	Shape	904.71	
Colors vs. Words	Color	861.00	.1729
	Word	813.31	
Colors vs. Shapes	Color	872.15	.0474
	Shape	998.11	
Words vs. Shapes	Word	808.60	.0006
	Shape	1019.60	

TABLE F16

Experiment 2: Newman-Keuls Test on Class Comparison Effect for the Dependent Variables of Recognition Latency (ms), Recognition Error (%), and SWAT Ratings*

Class Comparison	DxW	LxD	LxC	CxW	DxC	LxW	DxS	LxS	WxS	CxS
Recognition Latency	755	813	827	836	836	843	848	910	911	934

Class Comparison	DxW	LxD	LxW	LxC	DxC	CxW	DxS	LxS	WxS	CxS
Recognition Error	1.88	2.08	2.71	3.13	3.33	3.54	3.75	5.21	6.04	6.04

Class Comparison	LxD	DxW	DxC	LxW	LxC	DxS	LxS	CxW	WxS	CxS
SWAT Ratings	16.9	18.7	20.5	21.0	21.7	24.0	25.5	26.0	29.4	30.9

*Means with a common line are not significantly different at $p < .05$.

Table F17

Experiment 2: Summary of ANOVA results for the Letters versus Digits Class-Comparison; Dependent variable is SWAT Ratings

Source	df	MS	*F*	*p*
Subject	9	2733.0833		
ClaSize	2	44294.6159	48.50	.0001
ClaSize*Subj	18	913.3187		
PClas	1	10.8000	0.16	.7011
PClas*Subj	9	68.7456		
PType	1	471.6368	5.19	.0488
PType*Subj	9	90.9549		
ClaSize*PClas	2	74.4619	0.98	.3947
ClaSize*PClas*Subj	18	76.0415		
ClaSize*PType	2	569.5633	5.77	.0116
ClaSize*PType*Subj	18	98.6405		
PClas*PType	1	3.6401	0.05	.8226
PClas*PType*Subj	9	68.3444		
ClaSize*PClas*PType	2	13.1565	0.18	.8369
ClaSize*PClas*PType*Subj	18	73.1640		
Total	119			

Table F18

Experiment 2: Summary of ANOVA results for the Letters versus Colors Class-Comparison; Dependent variable is SWAT Ratings

Source	df	MS	*F*	*p*
Subject	9	3531.5314		
ClaSize	2	60062.6141	65.17	.0001
ClaSize*Subj	18	921.6359		
PClas	1	28.8610	0.10	.7623
PClas*Subj	9	296.8967		
PType	1	605.9260	1.88	.2033
PType*Subj	9	321.8540		
ClaSize*PClas	2	225.7436	2.21	.1390
ClaSize*PClas*Subj	18	102.3346		
ClaSize*PType	2	98.5963	0.33	.7206
ClaSize*PType*Subj	18	295.5027		
PClas*PType	1	693.8425	2.75	.1319
PClas*PType*Subj	9	252.6431		
ClaSize*PClas*PType	2	247.4703	1.65	.2197
ClaSize*PClas*PType*Subj	18	149.9193		
Total	119			

Table F19

Experiment 2: Summary of ANOVA results for the Letters versus Words Class-Comparison; Dependent variable is SWAT Ratings

Source	df	MS	*F*	*p*
Subject	9	2822.7601		
ClaSize	2	59025.7250	61.14	.0001
ClaSize*Subj	18	965.3416		
PClas	1	25.8541	0.22	.6534
PClas*Subj	9	119.8776		
PType	1	1730.5208	8.34	.0180
PType*Subj	9	207.5431		
ClaSize*PClas	2	68.5704	0.82	.4580
ClaSize*PClas*Subj	18	84.0447		
ClaSize*PType	2	312.2057	3.57	.0493
ClaSize*PType*Subj	18	87.3411		
PClas*PType	1	34.1333	0.19	.6754
PClas*PType*Subj	9	182.3438		
ClaSize*PClas*PType	2	281.5471	1.55	.2394
ClaSize*PClas*PType*Subj	18	181.7039		
Total	119			

Table F20

Experiment 2: Summary of ANOVA results for the Letters versus Shapes Class-Comparison; Dependent variable is SWAT Ratings

Source	df	MS	*F*	*p*
Subject	9	3586.7062		
ClaSize	2	71628.2214	59.22	.0001
ClaSize*Subj	18	1209.4298		
PClas	1	701.0750	5.48	.0439
PClas*Subj	9	127.8553		
PType	1	502.4567	0.83	.3857
PType*Subj	9	604.5043		
ClaSize*PClas	2	550.1170	2.46	.1137
ClaSize*PClas*Subj	18	223.7292		
ClaSize*PType	2	1.8751	0.01	.9910
ClaSize*PType*Subj	18	207.5396		
PClas*PType	1	177.7550	0.79	.3984
PClas*PType*Subj	9	226.1943		
ClaSize*PClas*PType	2	47.1115	0.13	.8819
ClaSize*PClas*PType*Subj	18	372.1992		
Total	119			

Table F21

Experiment 2: Summary of ANOVA results for the Digits versus Colors Class-Comparison; Dependent variable is SWAT Ratings

Source	df	MS	*F*	*p*
Subject	9	3536.7077		
ClaSize	2	55541.4417	44.45	.0001
ClaSize*Subj	18	1249.4417		
PClas	1	26.4141	0.22	.6535
PClas*Subj	9	122.5827		
PType	1	104.9070	1.06	.3310
PType*Subj	9	99.3753		
ClaSize*PClas	2	184.9576	1.61	.2267
ClaSize*PClas*Subj	18	114.6147		
ClaSize*PType	2	306.8676	1.50	.2492
ClaSize*PType*Subj	18	204.2269		
PClas*PType	1	213.8670	1.45	.2595
PClas*PType*Subj	9	147.6389		
ClaSize*PClas*PType	2	495.0294	3.13	.0683
ClaSize*PClas*PType*Subj	18	158.2670		
Total	119			

Table F22

Experiment 2: Summary of ANOVA results for the Digits versus Words Class-Comparison; Dependent variable is SWAT Ratings

Source	df	MS	*F*	*p*
Subject	9	2394.9378		
ClaSize	2	51994.0836	50.32	.0001
ClaSize*Subj	18	1033.2240		
PClas	1	405.5363	3.20	.1071
PClas*Subj	9	126.5570		
PType	1	3.1041	0.02	.9021
PType*Subj	9	194.0800		
ClaSize*PClas	2	114.7763	1.24	.3125
ClaSize*PClas*Subj	18	92.4320		
ClaSize*PType	2	159.9436	1.02	.3807
ClaSize*PType*Subj	18	156.8923		
PClas*PType	1	151.4253	1.30	.2836
PClas*PType*Subj	9	116.4529		
ClaSize*PClas*PType	2	256.4491	3.74	.0438
ClaSize*PClas*PType*Subj	18	68.5856		
Total	119			

Table F23

Experiment 2: Summary of ANOVA results for the Digits versus Shapes Class-Comparison; Dependent variable is SWAT Ratings

Source	df	MS	*F*	*p*
Subject	9	2490.6114		
ClaSize	2	55872.0016	87.00	.0001
ClaSize*Subj	18	642.1835		
PClas	1	3199.7177	8.97	.0151
PClas*Subj	9	356.73189		
PType	1	138.1380	0.35	.5692
PType*Subj	9	395.8503		
ClaSize*PClas	2	156.1508	1.42	.2671
ClaSize*PClas*Subj	18	109.8226		
ClaSize*PType	2	53.7309	0.34	.7131
ClaSize*PType*Subj	18	155.9449		
PClas*PType	1	47.1880	0.63	.4485
PClas*PType*Subj	9	75.1427		
ClaSize*PClas*PType	2	3.7101	0.02	.9778
ClaSize*PClas*PType*Subj	18	165.0850		
Total	119			

Table F24

Experiment 2: Summary of ANOVA results for the Colors versus Words Class-Comparison; Dependent variable is SWAT Ratings

Source	df	MS	*F*	*p*
Subject	9	5022.7664		
ClaSize	2	88782.8703	49.11	.0001
ClaSize*Subj	18	1807.9756		
PClas	1	43.9230	0.50	.4964
PClas*Subj	9	87.4181		
PType	1	168.0333	0.73	.4138
PType*Subj	9	228.8902		
ClaSize*PClas	2	60.3796	0.47	.6327
ClaSize*PClas*Subj	18	128.5959		
ClaSize*PType	2	570.5296	2.66	.0971
ClaSize*PType*Subj	18	214.2744		
PClas*PType	1	339.6968	2.67	.1368
PClas*PType*Subj	9	127.3062		
ClaSize*PClas*PType	2	207.5344	1.10	.3549
ClaSize*PClas*PType*Subj	18	189.0519		
Total	119			

Table F25

Experiment 2: Summary of ANOVA results for the Colors versus Shapes Class-Comparison; Dependent variable is SWAT Ratings

Source	df	MS	*F*	*p*
Subject	9	5977.2598		
ClaSize	2	92486.0974	70.43	.0001
ClaSize*Subj	18	1313.2299		
PClas	1	800.8333	4.90	.0542
PClas*Subj	9	163.5117		
PType	1	916.8741	1.99	.1915
PType*Subj	9	459.8057		
ClaSize*PClas	2	95.8604	0.34	.7192
ClaSize*PClas*Subj	18	285.5816		
ClaSize*PType	2	292.1090	1.69	.2124
ClaSize*PType*Subj	18	172.8031		
PClas*PType	1	299.5680	1.34	.2766
PClas*PType*Subj	9	223.3712		
ClaSize*PClas*PType	2	55.9734	0.28	.7584
ClaSize*PClas*PType*Subj	18	199.2642		
Total	119			

Table F26

Experiment 2: Summary of ANOVA results for the Words versus Shapes Class-Comparison; Dependent variable is SWAT Ratings

Source	df	MS	*F*	*p*
Subject	9	5282.3957		
ClaSize	2	93384.3250	112.15	.0001
ClaSize*Subj	18	832.6404		
PClas	1	1394.0083	4.40	.0654
PClas*Subj	9	316.9912		
PType	1	84.0013	1.67	.2288
PType*Subj	9	50.3877		
ClaSize*PClas	2	99.2536	0.77	.4759
ClaSize*PClas*Subj	18	128.2172		
ClaSize*PType	2	80.6893	0.31	.7358
ClaSize*PType*Subj	18	258.5673		
PClas*PType	1	24.9341	0.05	.8241
PClas*PType*Subj	9	476.3129		
ClaSize*PClas*PType	2	89.4691	0.28	.7556
ClaSize*PClas*PType*Subj	18	314.3199		
Total	119			

TABLE F27

Experiment 2: Probe Class SWAT Ratings and Associated p-value for each Class Comparison

Class Comparison	Probe Class	SWAT Ratings	p-value for Probe Class Effect
Letters vs. Digits	Letter	16.69	.7011
	Digit	17.16	
Letters vs. Colors	Letter	21.29	.7623
	Color	22.18	
Letters vs. Words	Letter	20.99	.6534
	Word	21.05	
Letters vs. Shapes	Letter	24.53	.0439
	Shape	26.47	
Digits vs. Colors	Digit	20.90	.6535
	Color	19.99	
Digits vs. Words	Digit	19.73	.1071
	Word	17.67	
Digits vs. Shapes	Digit	21.66	.0151
	Shape	26.49	
Colors vs. Words	Color	26.12	.4964
	Word	25.86	
Colors vs. Shapes	Color	29.38	.0542
	Shape	32.49	
Words vs. Shapes	Word	28.02	.0654
	Shape	30.87	

Table F28

Experiment 2: Pearson's Product Moment correlation coefficients among the variables of Memory-Set Size, Recognition Latency, and SWAT Ratings*

	Memory-Set Size	Recognition Latency	SWAT Ratings
Memory-Set Size	. .	0.2412 0.0001	0.6966 0.0001
Recognition Latency	. .	. .	0.3232 0.0001

*All correlation coefficients are significant beyond .0001.

VITA

Kay Chuan Tan

Kay Tan was born in the Republic of Singapore on the 27^{th} of September, 1960. He earned a B.S. in Psychology in 1984, and an M.S. in Industrial Engineering and Operations Research (Human Factors) in 1987, both from the University of Massachusetts at Amherst. His thesis topic was "Type of Highlighting and Search Strategy Considerations in the Use of Computer-Generated Displays."

Kay served as a research assistant during periods of 1985 to 1987 in the Human Performance Laboratory at the University of Massachusetts. During the period April, 1986 to December, 1986, he was a Human Factors Research Consultant with Wang Laboratories, Lowell, MA, where he was involved with the evaluation of computer design specifications and interface rapid prototyping.

Kay began doctoral work at Virginia Polytechnic Institute and State University in 1987. He spent nine months with Drs. Deborah Hix and Rex Hartson from the Computer Science Department in developing an evaluation form and procedure for the assessment of User Interface Management Systems. Under the tutelage of Dr. Robert Beaton, he began dissertation work on Short-Term Memory and Cognitive Workload in January, 1989.

Kay is a student member of the Human Factors Society, the Institute of Industrial Engineers, and the Association of Computing Machinery. He received membership honors from Psi Chi and Alpha Pi Mu. To date, Kay has authored or co-authored three technical reports, four conference papers, and one journal paper. He completed all doctoral requirements in Industrial Engineering and Operations Research (Human Factors) in June, 1990.