AN ABSTRACT OF THE THESIS OF

Carlos Andres Vasquez Aguirre for the degree of Master of Science in Industrial Engineering presented on October 29, 2004.

Title: A Preliminary Study for Task Management Environment External Validation: Correlation Between Continuous, Discrete and Mixed Scenarios

Abstract approved:

Redacted for Privacy

Kenneth H. Funk, II

Concurrent Task Management (CTM) can be defined as the process that human operators of complex systems perform to allocate their attention among multiple, simultaneous tasks. A challenge of CTM research is to create an experimental environment which is complex enough to model a real world multitasking situation. Also, this system has to be operable by participants without special qualification or extensive training. The answer to this challenge is the Task Management Environment (TME), which is a low fidelity multitasking management program. To further explore more CTM theories, it is necessary to make an external validation of the TME. In order to perform this validation there is a previous step that is necessary in order to find the optimal TME setup to validate. The objective of this research is to look for a possible relationship in performance between continuous, discrete and mixed task scenarios. The study tested 75 participants in three different TME scenarios (continuous, discrete and mixed). The results showed that there is not a significant correlation between continuous and discrete Scenarios, and there are small correlations between continuous and mixed scenarios and discrete and mixed scenarios. These results suggest that the optimal scenario to be tested for the validation study is a mixed scenario.

A Preliminary Study for Task Management Environment External Validation: Correlation between Continuous, Discrete and Mixed Scenarios

by

Carlos Andres Vasquez Aguirre

A THESIS

Submitted to

Oregon State University

in partial fulfillment of
the requirements for the
degree of

Master of Science

Presented October 29, 2004
Commencement June 2005

Master of Science thesis of Carlos Andres Vasquez Aguirre presented on October 29, 2004

APPROVED:

Redacted for Privacy

Major Professor, representing Industrial Engineering

Redacted for Privacy

Head of the Department of Industrial and Manufacturing Engineering

Redacted for Privacy

Dean of Graduate School

I understand that my thesis will become part of the permanent collection of Oregon State University libraries. My signature below authorizes release of my thesis to any reader upon request.

Redacted for Privacy

Carlos Andres Vasquez Aguirre, Author

ACKNOWLEDGMENTS

During this stage of my life I have received support from many professors, friends and family. This has been a great experience of learning the value of knowledge, friendship and love. I would like to express my most sincere acknowledgement to all.

I would like to thank my Mom, Patricia, for her constant sentimental and economical support through my life here in Corvallis. Without her, I probably could not have made it to this point; her constant communication was of value to cheer me up and to know that somebody was always there for me. I would like to thank my Dad, Anibal, for his support and encouragement for me to be a successful person in life and always looking forward to become a better person. I would like to thank my brother Jose, for his advice about the university life at OSU and for sharing with me his experiences here at Corvallis. And finally, I would like to thank my Grandma Nicol, for her prayers and unconditional love, and my Aunt Consue, for the advise on how to be successful in life and for pushing me to be somebody in life.

I would like to thank my major professor, Dr. Ken Funk, for his constant advice through my master's studies and especially for his guidance in this project. Thank you for all the time that you have spent during our continuous meetings, your time spent in reading my drafts to give me constructive feedback and also for helping me to choose and schedule my program of study with the purpose of being most successful.

Finally, I would like to thank all my friends, particularly the Association of Latin American Students (ALAS), which has been my second family through this stage of my life. Thank you guys for supporting me through the good and bad times. *Gracias por ser mis amiguitos!!*

TABLE OF CONTENTS

TABLE OF CONTENTS (Continued)

LIST OF FIGURES

LIST OF FIGURES (Continued)

LIST OF TABLES

LIST OF APPENDICES

CHAPTER 1 INTRODUCTION

Concurrent Task Management (CTM) is the process by which operators of complex systems such as an airplane cockpit, or an operating room; selectively attend to multiple tasks at the same time in order to complete an assignment satisfactorily. This subject has been researched for many years. In order to expand the horizons of continued investigation in the field and the theories of CTM, it is necessary to identify a simple tool that can measure CTM performance.

One possible tool to measure CTM performance is called the Task Management Environment (TME). TME is a computer program that simulates an abstract system, composed of simple dynamic subsystems that represent subsystems in a real world scenario. An important question to consider is whether or not this program really measures concurrent task management performance. To answer this question, it is necessary to make an external validation of the TME comparing a proven concurrent task management measurement device, such as a flight simulator. It will then be compared to the TME and the performance results of the two devices will be compared, to determine if there is a positive correlation.

Before performing this validation, it is necessary to undertake a preliminary study of TME to determine how to configure the TME and identify the best-fit scenario to be run for the validation study. In order to find the most appropriate, the present research was done.

This document begins with Chapter 2, introducing the Concept of CTM, with a presentation of a literature review of attention, divided attention, selective and focused attention, multitasking/time-sharing tasks, workload, strategic workload management and multiple resource theory and their relationship to CTM. There is a sample of some research studies relevant to CTM. Following this section, there is a summary of CTM findings from the literature review of what we do know and do not know about CTM. The final portion of chapter 2 is a summary of the research performed using the Task Management Environment and a statement of justification for this thesis study.

The research objectives of this study are contained in Chapter 3. These are: a correlations study of continuous, discrete and mixed TME scenarios performance and a study of the strategies participants used in continuous scenarios.

Chapter 4 explains the TME function and elements. Also, it describes the research methodology in detail, including the experimental procedure and data analysis procedures.

Chapter 5 describes the results of analysis of the data obtained.

Chapter 6 interprets and discusses the results of this research and presents the limitations and suggestions for future research.

CHAPTER 2 LITERATURE REVIEW

2.1. Introduction

This chapter presents a summary of a literature review of the most important topics related to Concurrent Task Management (CTM), multitasking theory and other associated topics. The first section contains definitions regarding attention, divided attention, selective and focused attention, dual task, multitasking, workload, strategic workload management, multiple resource theory, and the relationship of these theories to Concurrent Task Management (CTM). The second section contains examples of research relevant to CTM, which includes laboratory, driving and aviation research. Following this section, there is a summary of CTM findings from the literature, which is divided into what we "know" and open questions about what we "do not" know yet about CTM. Finally the last part of this chapter is an introduction to the Task Management Environment (TME). This section contains an introduction to the TME software and TME research along with a summary of TME findings. The conclusion of this chapter is an open question about the justification for the problem statement, which is established on the next chapter.

2.2. Definitions of Concurrent Task Management and Related Topics

This section contains definitions of divided attention, selective/focused attention, multitasking/time-sharing, workload, strategic workload management and multiple resource theory, and the interrelations among these topics with CTM.

Concurrent Task Management (CTM) can be defined as the process that human operators of complex systems perform to allocate their attention among multiple and concurrent tasks (Nicolalde, 2003). CTM is a vulnerable point because the attention process is a severely limited resource (Dismukes et al 2001). CTM is a relatively new theory, but research related to CTM goes back to the initial studies on attention and cognitive resource management.

The early studies of limitations of cognitive resources were centered on attention. Specifically divided attention is directly related to CTM. Divided attention studies the correct distribution of attention among the assigned activities emphasizing the switching of attention among those activities. Divided attention can also be defined as parallel processing between two or more channels of information or stimuli (Goesh, 1990). Attention on the other hand, deals with the theory of a cognitive tunnel and the limitation of information flow. Broadbent (1958) used the analogy of a bottleneck that allows information to flow only one piece at a time and act as a selective attention filter, which allows information to flow from one channel to another. Treistan (1968) studied attenuation and selective attention. Selective attention recognizes the characteristics of the message, chooses the information to attend to, and finally reacts to the obtained information. Deutsch and Deutsch (1963) concluded that the retrieval of information depends on the degree of importance. Norman (1968) did a complementary study of this theory. His findings depend on the degree of activation of information according to the strength level of the input variable.

Selective attention is the condition when a person chooses to attend to cues that stand out instead of paying attention to all cues (Broadbent, 1982). In a particular mission, the selection of the right cues is a task of setting goals, and attending to the right cues at any given time. Therefore, selective attention can be included as a component of CTM. Differing somewhat from selective attention, but also considered a part of CTM, is focused attention.

Focused attention occurs when the person concentrates on only one source of information or cue signal. There is a difference between selective attention and focused attention, according to the aviation literature. Selective attention serially determines what relevant information in the environment needs to be processed. Focused attention refers to the ability to process only the necessary information and filters out what is unnecessary (Prinzel, 2004).

Focused attention is utilized when there are several sources of information and the person chooses to process these selectively. A misallocation of attention or focusing attention on an inappropriate cue could be defined as attending to one task when another task has a higher priority, consistent with a rational task prioritization strategy. One of the most common examples of studies in focused attention is the dichotic listening task. This study makes the person listen to two different sound stimuli simultaneously and try to monitor or shadow information from either one of the stimuli (Woodworth, 1938).

Later studies of human performance became more complex with dual task studies and multitasking studies. Those studies researched how people recognize or handle two or more different tasks concurrently. The most common studies consist of doing a cognitive task at the same time that the person is performing a psychomotor task (e.g. a tracking task or reaction time task). More realistic studies investigate multitasking/time-sharing, which can be defined as the execution of two or more tasks simultaneously or in a very short period of time. An example of multiple tasks in flying an airplane is the common situation when a crew member must interleave (alternate) the steps of two or even more procedures simultaneously, for example, keeping track of position in relation to the taxiing clearance while watching for conflicting traffic. Other examples are reading a checklist, monitoring taxiing progress and responding to radio calls at the same time.

Workload is a way to explain multitasking failures and is defined in terms of the human's limited processing resources and the ability to perform concurrent tasks that are either required, or potentially required (Wickens, 1992). Workload is defined by the relationship between resource supply and task demand or by the specification of the amount of information processing capacity that is used for task performance (Waard,1996).

Multiple Resource Theory (MRT) describes the utilization of processing capacity in order to perform several tasks. The MRT structure concept forecasts the effects of performance according to how multiple tasks are executed. There are limited resources (attention) that can be allocated to a specific task in order to complete a set of tasks. If in a given situation a secondary task needs to be performed and the primary task has the allocation of full resources, the primary task deteriorates as a result of the attention drawn away from it to the secondary task (Wickens, 1980, 1992).

Strategic Workload Management (SWM) explains the process of attention and task management where a human manages division of workload based on the difficulty of the tasks. Also SWM assures that the workload is at an ideal level; not too high or too low. SWM can be defined as the task management strategy of deciding what to perform or not to perform. SWM is closely related to the decision of when to perform a task and to the currently experienced level of workload (Wickens, 1994).

In spite of the different terminology of these studies and their findings, the studies point to a common subject, Concurrent Task Management. The simplest studies related to CTM were the cognitive tunnel and allocation of attention between tasks studies. These were subdivided into studies of attention in different terms (focused and selective). Later on, research was conducted in the allocation of resources among several tasks at the same time. This is called multitasking. The most recent research studies address the limitation of resources allocated to tasks with the theories of workload and multiple resource theory. Finally, the study of strategic workload management explains the allocation of resources among different tasks depending on the status of the task and other factors.

The following section contains examples of research relevant to CTM. It includes examples of general, driving and aviation research. At the end of this section there is a summary of CTM findings from the literature review, which is divided into what we “know” and some open questions about what we “do not” yet know about CTM.

2.3. Research Relevant to CTM

This section is divided into different CTM research case studies. The section includes general (not domain specific), driving and aviation case studies.

2.3.1. General Studies

Humphrey and Kramer (1999), studied and analyzed the age-related differences in the use of visual environment to facilitate the processing of task relevant stimuli. The experiment consisted of showing to the participants different stimuli in a display. Then the participants were asked to recall the stimuli. The results showed that older adults demonstrated greater sensibility in grouping by proximity and similarity than did the younger adults. This helped to make concurrent task management in a closer display with a first hand information compared to a display with information split in different areas.

Bondar (2002), researched how older and younger adults allocate their mental resources in dual task situations that involve sensorimotor and cognitive components. The results show that older participants first shift all of their attention to the cognitive task. Also, older participants always protected their balance and then tried to attend to the cognitive task. But this capacity is limited when resource demand of sensorimotor tasks increases.

2.3.2 Driving

Strayer and Johnson (2001), studied the effect of cellular phone conversation on driving performance. The experiment consisted of a pursuit tracking task on a computer display, which turned to red or green in order to simulate a traffic signal, while the person was engaged in a conversation by cell phone using a hand held or hands free device. The results show that persons whom were engaged in a cellular phone conversation failed more frequently to detect simulated traffic signs. The investigators concluded that a cell phone conversation distracts the person from paying attention to the road due to interference with the central attentional processes and deteriorates their driving performance.

Radeborg et al (1999), studied the kind of concurrent activity that affects a person's working memory. This study uses a low fidelity driving simulator. The primary task was to drive on a sinuous (curvy) road with two different difficulty levels (easy and difficult). The secondary task was to identify the meaning of a sentence or recall words from a sentence. The results of the study showed that information processing is affected by the intrinsic demands of driving. However, the difficulty of the driving task does not seem to be a factor. This means that the type of driving scenario did not affect the working memory; although the type of information or conversation carried on could affect both driving performance and verbal judgment.

Summala (1996), researched the differences of driving abilities between novice and expert drivers. The experiment was performed in a real world scenario on the road with different display locations. The drivers were asked to provide information obtained from the different displays. The research showed that novice drivers used more their vision resources to focus on the road and used the mirrors to look at adjacent lines and rear conditions. On the other hand, expert drivers shared attention between driving and other activities. Therefore, drivers may cope with dual tasks by dividing attention within the visual field. As a result, the attention task showed a relationship between task location and driving experience. Also, results confirmed the hypothesis of Mourant and Rowell (1972), which indicates that novices need foveal vision for lane keeping. But, with more practice they learn to manage it with more peripheral vision as expert drivers do. The research confirmed that practice and experience are an important factor that influences the performance of concurrent tasks, in this case driving tasks.

Dingus (1997), studied the effect of selected in-vehicle route guidance systems on the attentional demands of driving. The experiment was performed in a high fidelity driving simulator. Participants were tested on different route guidance systems. The results show that better performance in driving and faster reaction time to external events was produced by the use of audio guidance and the poorest performance and slowest response time was produced by a paper map. Audio systems permitted the driver to pay more attention to the road and respond faster to unusual events. The heads up display reduced the need for taking eyes off the road when obtaining information from the display and allowed participants to respond faster. On the contrary, head down displays required turning the head or glancing at the display at the same time that the driver is performing a driving task.

Sodhi et al (2002), researched and measured the distraction effect of many in-vehicle devices and different tasks on driving performance. The participants were asked to drive on a preseleted two-lane road, while at the same time, they were being asked to do different tasks (visual and cognitive). The results show that there is a difference between the off-road and on-road glance times. The reason might be that both the radio and the rear-view mirror tasks require significantly more data collection compared with reading a sign on the road, which is located on the same visual field. For the cognitive task, drivers did not scan the road as much as they would otherwise. This implies that cognitive tasks increase distraction from driving performance.

2.3.3. Aviation

Crosby and Parkinson (1979), measured the performance of instructor pilots vs. student pilots in a dual task paradigm combining a ground controlled approach as the primary task and a memory search as the subsidiary task. The students were tested during the middle phase of their training and then after completion of their training. The results showed that memory search performance improves between the first phase of training and the final phase of training. This findings implies that training does possibly influence dual task management and is a potential measure of flight proficiency.

Milke et al (1999), researched the relationship between age, cognition, and pilot performance. Subjects performed the study with a computer neuropsychological test battery. The result indicates that in the case of divided attention tasks, significant age vs. condition (single or multiple task) interactions are correlated. Concurrent task performance of older people was differentially affected, depending on the workload demand.

Tsang (1986), researched the relationship between the training undergone by pilots and their increased performance on time-sharing with dual tasks. The test consisted of professional pilots and college students performing a tracking task and a transformation task simultaneously, similar to a dual task test. The results indicated that pilots' time-sharing performance was not much better the students' in dual task performance. However, the improvement of most participants' performance throughout the experiment indicates that time sharing performance can be improved with training and practice.

Lamoreux (1999), studied the influence of aircraft proximity data on the mental workload of air traffic controllers. The complexity of relationships between various aircraft and their motion were also studied. In addition, this research studied how the proximity of multiple aircraft alters the mental workload of air traffic controllers and also why only a small part of these alterations may be attributable to the number of aircraft in a sector. It was concluded that workload might vary for the behavior or complexity of the tasks, rather than for the quantity of tasks.

Funk (1991), introduced and defined the term Cockpit Task Management (CTM). CTM is the process by which pilots selectively attend to tasks in such way as to achieve their mission goal. CTM determines which concurrent tasks pilot(s) attend to at any given time.

Colvin (2000), studied the factors that affect task prioritization on the flight deck. The participants of this study were asked to use a flight simulator in two different scenarios with different events (malfunctions, procedures and Air Traffic Control call). Then, participants were interviewed to determine what factors influenced the task prioritization process. This was done using an intrusive method (during the simulator test) or retrospective method (after the simulator test by reviewing a video). The pilots' responses were categorized into 6 main factors: status, procedure, value, urgency, salience and effort. Later, a second study of the six main factors that influenced the task prioritization was researched. The results showed that the most important factors that influence pilot's response are status, procedure, and value.

Funk and Braune (1999), developed and evaluated an experimental computational aid called the Agenda Manager (AMgr) to facilitate agenda management as an extension of Cockpit Task Management. The test consisted of pilots performing a partial task in a flight simulator. The objective was to compare the usage of AMgr to a conventional monitoring and alert system. The results of this study showed that AMgr was a better tool in recognizing goal conflicts and solving those problems in a shorter amount of time. Also, the results showed that use of the AMgr improved pilot performance which could, in turn, improve flight safety. However, the implementation of this device is restricted by some FAA regulations and by technology constraints in certain types of aircraft.

Bishara (2002) conducted a study that suggested that the amount of training was related to the task management response. Also, training increased the ability to reduce task prioritization errors. A follow up study was performed to confirm the conclusion from the previous study. However, the results showed that training in CTM did not significally influence CTM performance. A question was then raised about whether or not CTM can be improved by practicing or training.

2.3.4. Summary of CTM findings and open questions

An interesting find in CTM is how the level of experience or expertise on any multiple task management situation (e.g. driving, flight) can influence the concurrent task performance. Other aspects that were found through CTM research studies were that:

1 Older age limits the capacity to manage multiple tasks and processing resources.
2 The use of different channels of information improves task management.
3 The interference of distractions with the central processes influences the performance of task management.
4 CTM is time-driven. This means that it is a process of scheduling or ordering tasks in a timely fashion.
5 Prioritization of tasks depends upon the characteristics of the task and its importance in any given scenario.
6 The workload level influences the performance of CTM. This level should be in an optimal level, not too high to produce stress or too low to produce boredom.

At the same time there are questions that are extracted from studies on CTM. Some examples are:

1 Can a subject be trained in CTM?
2 Can a subject be aided in CTM?
3 Does strategy affect CTM performance?
4 Does gender, age, or experience influence the performance of CTM?
5 Are status, procedure and value the most important factors that influence CTM? Are there any other factors?
6 Does distraction influence the lack of attention control and the inability to efficiently allocate resources?

In order to continue this line of research, it is desirable to create a tool to test some CTM theories and conduct more research. A common challenge with CTM research is creating an experimental environment that is complex enough to accurately model a real world multitasking situation (e.g. driving, flying an airplane or controlling air traffic); yet create a system that is operable by participants without special qualifications or extensive training. A possible solution to this challenge was the creation of TME.

2.4. The Task Management Environment (TME)

The Task Management Environment software is a low fidelity multitasking management program developed by Shakib (2002). The original idea of the TME was developed in a program called "Tarstad" (Persian for juggler). The program is based on the metaphor of a juggler spinning plates to simulate the activity of a multitasking operator such as an airplane pilot, car driver or an air traffic controller. Each task is represented as a plate that is spinning on a stick. The amount of attention that is being paid to a task is represented by the speed of the spinning plate. Less attention means the plate loses speed and starts to fall and vice versa. The degree of importance of the task is represented by the type of plate that is spinning (crystal, porcelain, glass, etc).

TME simulates an abstract multiple system composed of up to 15 simple, dynamic subsystems (Figure 2.1 shows only 8 subsystems). A subsystem is what the TME models to represent a task. Each subsystem has a variable that ranges from 0% to 100%, called status (S) and is represented by the height of the blue bar of each subsystem.

There are two types of subsystems, continuous and discrete. When a continuous subsystem is unattended, its status decreases at a constant rate until it reaches 0%. This is called the Deterioration Rate (DR). On the other hand, when the operator clicks the computer mouse on the subsystem's grey button, the status increases at a constant rate with a limit of 100%. This is called the Correction Rate (CR). When the operator releases the mouse button the status begins to decrease.

A discrete subsystem has a similar behavior, except its that status remains at 100% until a simulated failure event occurs. At that moment, its status decreases at a rate of DR until the operator clicks the grey button. The decrease temporary stops for a predetermined time while the subsystem button fades (is inactive). After a few seconds, the button reappears with a status of 100% and the status starts dropping with the same DR rate until the button is clicked again. This process continues for a predetermined number of cycles until status is restored to 100%. A new failure could occur anytime.

The importance of each subsystem is represented by the number located in the middle of the control button and is called the weight (W).

Scenario refers to the different type of setups of TME depending on the type of subsystems that are chosen to be tested (continuous, discrete or mixed).

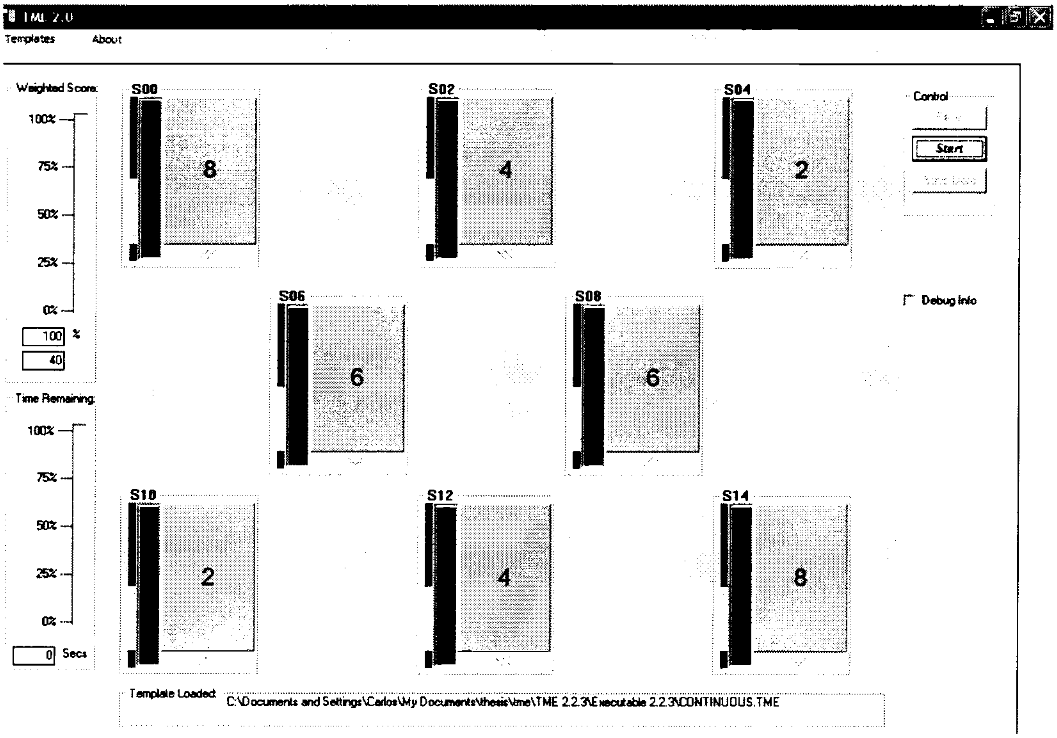

Figure 2.1 TME Interface

To understand more about the function and set up of TME, please refer to Chapter 4, section 4.2.2. Software used for the research, which includes explanations about type of scenarios, parameters, and examples of subsystems in the TME. Also in order to know more about how the CTM performance is measured, please refer to Chapter 4, section 4.3.1.2. Total Weighted Average Score.

The following are examples of the CTM research that has been performed with TME software:

Shakeri and Funk (2003), made a comparison between humans and mathematical search-based solutions in managing multiple concurrent tasks. The main objective in this research was to create a tool to measure human multitasking performance and to compare this with an optimal score. This was achieved by using a tabu search based heuristic method. The participants ran five different scenarios in TME with differences in Deviation Rates, Correction

Rates and Weights. The results showed that none of the participants could obtain an optimal score in any scenario. Also, participants tried to handle all the tasks, instead of developing a good strategy for task management. The participants only focused on getting a better score.

Nicolalde (2003), researched the relationship between cognitive abilities and CTM. The experiment consisted of testing the participants on different cognitive tests (reaction time, verbal IQ, etc). Also, the participants were trained and faced to two different scenarios in TME (easy and difficult). The results indicate that the correlation between cognitive test and TME results was low. This was interpreted to mean that CTM is a complex cognitive process that cannot be measured or explained by elementary cognitive abilities.

Chen and Funk (2003), developed a fuzzy model to show how people prioritize tasks based on their level of importance, the status of the task, and the urgency to perform. Six fuzzy models were created in order to analyze data obtained in TME. The six models were: 1) Random model that attends randomly to any task on the scenario; 2) Status (S) ,which chooses the task with the lowest status for attention next ;3) Status and urgency (SU) ,that chooses to attend to the task with the lowest status and urgency, no matter the weight or degree of importance of the subsystem; 4) Status and importance (SI) ,which considers the importance and the status of the subsystem; 5) 5major strategy, which focuses on the five most important weighted subsystems, and 6) 4major strategy, which focused on the four most important weight subsystems. The models S, SU exhibited performance most similar to to task of human participants, suggesting that those factors are considered in human CTM and SI did influence participant CTM strategies.

It is important to validate this tool in order to explore more Concurrent Task Management theories and to generalize any results obtained with the TME. In order to do the external validation of TME, it is necessary to perform a comparison study between a valid tool of CTM such as a flight simulator or real world situation (e.g. real flying or driving) and the TME. A flight simulator is a tool approved by the FAA as a training device. Thus, it is accepted as a validated measurement tool of CTM. Unfortunately, this device is complex to manage, expensive and not accessible to every researcher. Also, a real world situation has many time, safety, and cost constraints.

In order to design the TME external validation study, there are a few steps that have to be performed before the actual study. The first step is to create set up which will be compared to the simulator. In order to create this set up (scenario), it is necessary to make a correlation analysis between different types of set-ups in TME. This is necessary to investigate the type of scenario used to measure task management. For this reason, the primary objective of this thesis was to do a preliminary study of different kinds of scenarios (continuous, discrete and mixed) and to investigate the relationship among them. The secondary objective is to make an strategy analysis of a continuous scenario or continuous setup based on one of the questions unanswered about the CTM. Specifically, what type of strategy influences CTM?

CHAPTER 3 PROBLEM STATEMENT

In response to several of the unanswered questions raised in the literature review, and due to the lack of confidence in the Task Management Environment (TME) to accurately measure concurrent task management performance, it is important to make an external validation of TME as a tool to measure CTM. The first step in this validation process is to create an appropriate setup (scenario) capable of measuring multitasking performance. This setup will be compared to a simulator or a real world multitasking situation. In order to create this setup, it is necessary to establish the main relations or differences between continuous and discrete tasks in the TME. Also, it is necessary to establish if the setup influences in the performance of CTM in order to create a setup more similar to the real world. The research described below is the beginning of a long process to accomplish the goal of validation of the TME.

The main objective of this research is to determine whether or not there is a correlation between Concurrent Task Management performance in all continuous, all discrete or mixed scenarios. The objective is to determine if the relationship (correlation) does indeed exist and if so, how strong it is. This finding would indicate whether or not it matters which scenario (continuous, discrete or mixed) is used for the external validation experiment. On the other hand, if the relationship is minimal or does not exist at all, it is necessary to identify which setup would best suit the external validation.

Additionally, the collection of data to meet this objective creates an opportunity for research in different task management strategies used by the participants.

Therefore the second objective of this research is to determine whether or not there is a common strategy used by people when performing Concurrent Task Management in continuous scenarios.

CHAPTER 4 METHODOLOGY

The main purpose of this chapter is to describe the process of the experiment, the participants, the experimental design, and the process for the analysis of the data. Also, this section explains the creation of three different TME scenarios (continuous, discrete and mixed), based on tasks extracted from real world scenarios.

4.1 Participants

This study used a total of 81 participants (6 for the pilot study and 75 for the study) recruited from the population of Oregon State University students, faculty and staff members. Participation was completely voluntary and none received any reward or compensation. The participants were not restricted by gender, age, student status or ethnic group.

4.1.1. Recruitment

Participants were recruited by bulletin board announcements, fliers, e-mails to several campus associations and by word of mouth at OSU during winter and spring terms 2004.

4.2 Apparatus

This section describes the equipment (hardware and software) used in this research project.

4.2.1. Hardware

The computer used for running the experiment was a Dell GX1p Intel Pentium III with 512 MB of RAM, 498 MHz processor and Microsoft Windows XP operating system, video sonic monitor (Dell P991 in All-In-Wonder 128PCI) with screen resolution of 1024 by 768 pixels and a standard optical mouse (Logitech).

4.2.2. Software

The software used in this research is called the Task Management Environment. The following section explains the basics of TME, the type of scenarios, and how the scenarios were set up in the program.

4.2.2.1. *Introduction*

The TME is a software program that simulates a multitasking experimental environment. The TME is written in Microsoft Visual Basic 6.0 and runs on the Windows Operating System. Each TME scenario used in this study was composed of eight subsystems of either kind, continuous or discrete. Each subsystem interface is composed of five elements which are the importance number, located in the middle of the button; the status bar, that ranges from 0% to 100% and is represented by a blue vertical bar in the interface; the color zone bar (green, yellow, red from top to bottom); the control button; and the button window under the control button that represents the number of clicks until the recovery of discrete subsystems. A single subsystem interface is illustrated in Figure 4.1

The version used to run the experiment was TME 2v2, developed by Jorge Moncayo (2003).

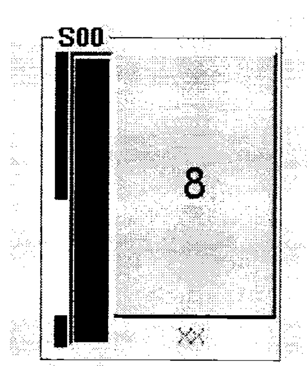

Figure 4.1 Single Subsystem Interface

4.2.2.2. *Scenarios Set up*

TME 2v2 has the capacity to create different types of scenarios by accessing a simple interface of the program (Figure 4.2). This interface has the capacity to modify different parameters of the program such as:

1 Type of behavior. This is the option to choose behaviors (continuous, discrete or general as described below).
2 Number of subsystems for the scenario.
3 Correction Rate (CR) for each subsystem. This is the rate at which a subsystem recovers or rises its status to 100%.
4 Deviation Rate (DR) for each subsystem. This is the rate at which the system decreases its status to 0%.
5 Weight for each subsystem. This is the degree of importance of each subsystem.
6 Active subsystem. This option enables a subsystem to be part of a scenario.
7 Time to run the experiment for each system (Time Span). This is the time allowed to run the experiment .
8 Time to fail. This is the time elapsed for each subsystem before it starts to fail or drop from the 100% status.
9 Discrete maximum count. This is the number of clicks that a participant needs in order to restore discrete subsystem's status to 100%.

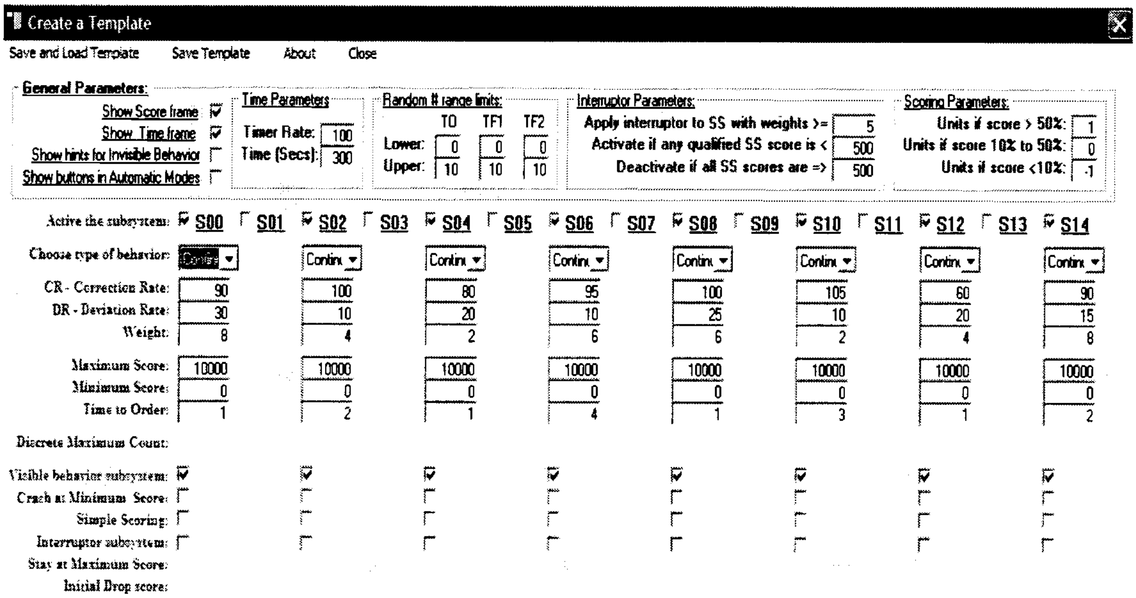

Figure 4.2 Interface to modify TME parameters

4.2.2.2.1. Scenarios for the experiment

For this experiment, three types of scenarios were created. The first one was an all continuous scenario consisting of eight continuous subsystems. The discrete scenario was composed of eight discrete subsystems. And finally, the mixed scenario was composed of a mixture of four subsystems from the continuous scenario (Subsystems 00, 04, 08 and 12) and four subsystems from the discrete scenario (Subsystems 02,06,10 and 14). The numbers in the parenthesis indicates the label position for each of the subsystems in the interface and is shown in the top left corner of each subsystem interface (Figure 4.1).

The parameters to set up the different scenarios in the TME for the experiment are shown in the Table 4.1

Table 4.1 Parameters for TME setup

	S00	S02	S04	S06	S08	S10	S12	S14
CONTINUOUS								
DR	90	100	80	95	100	105	60	90
CR	30	10	20	10	25	10	20	15
WEIGHT	***8***	***4***	***2***	***6***	***6***	***2***	***4***	***8***
DISCRETE								
DR	90	90	80	70	100	110	60	80
CR	55	5	35	20	60	15	65	40
WEIGHT	***8***	***4***	***2***	***6***	***6***	***2***	***4***	***8***
D. MAX. COUNT	5	3	5	2	5	2	4	3
MIXED								
DR	90	90	80	70	100	110	60	80
CD	30	5	20	20	25	15	20	40
WEIGHT	***8***	***4***	***2***	***6***	***6***	***2***	***4***	***8***
D. MAX. COUNT	*	3	*	2	*	2	*	3

4.2.2.2.2 *Continuous Subsystems*

A TME continuous subsystem is a representation or model of a real world subsystem that requires constant attention in order to keep it at a satisfactory level. This type of subsystem has a constant change of status. Figure 4.3 illustrates how continuous subsystems behave in TME. When the control button is held down the status bar recovers at the correction rate (CR); on the other hand, if the control button is not held, the status bar drops with some deviation rate (DR). Time holding (Th) in the figure represents the time the participant holds down the control button with the mouse.

A good example of a continuous task in the real world is steering a car. The driver has to pay constant attention to the road and manipulate the steering wheel to correct any deviation of the vehicle from the heading line.

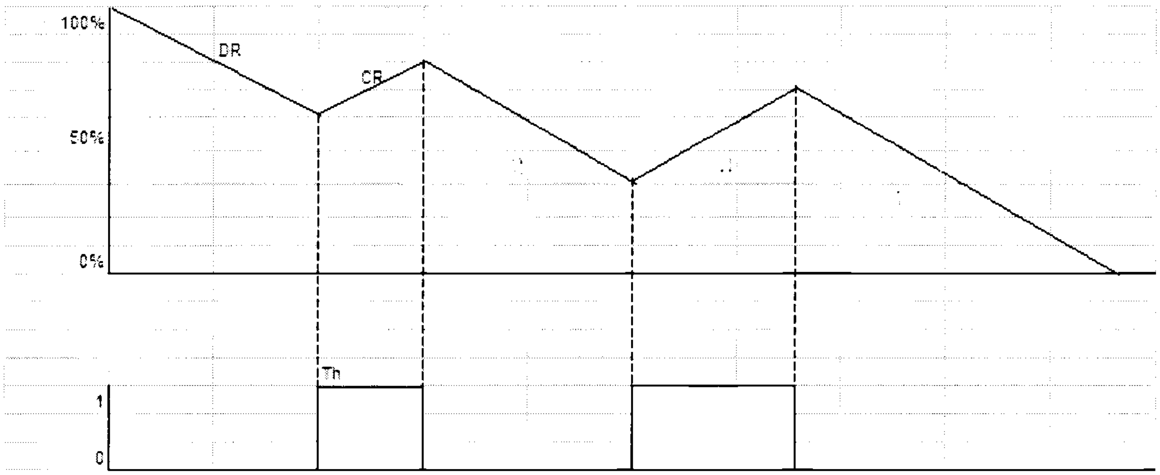

Figure 4.3 Continuous Subsystem Behavior

4.2.2.2.3. *Discrete Subsystems*

A TME discrete subsystem is a representation of a type of real world system which has a random change of state (a failure) that requires a series of discrete actions in order to restore it to normal. Figure 4.4 illustrates how discrete subsystems behave in TME. The subsystem stays at 100 % until a failure occurs (tf), the status bar drops with some deviation rate (DR) and is held when the participant makes a click of (pushes and releases) the control button. After a random time (td), the status of the subsystem drops again and the status bar starts to drop again until the participant clicks the control button. This cycle continues for a predetermined number of times until the system recovers again to 100% where it remains until a new failure occurs.

An example of this type of task is performing maintenance on a machine which requires a series of steps to bring it back to normal function. The technician has to perform an evaluation and figure out what is wrong with the machine. Then, the technician has to get new parts for the machine, repair or exchange parts, and finally restart the machine to return it normal functioning.

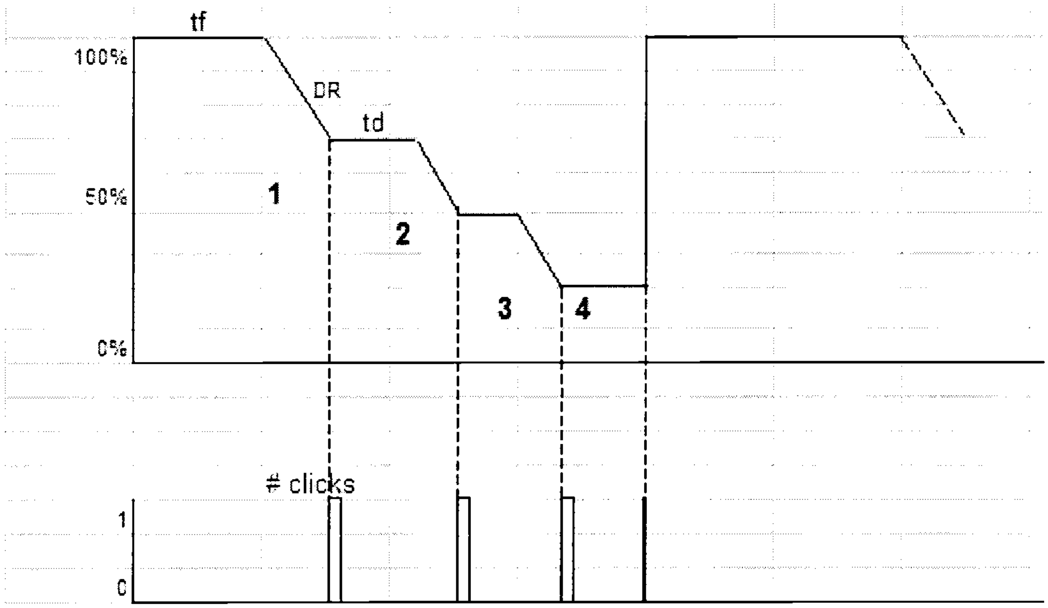

Figure 4.4 Discrete Subsystem Behavior

4.3 Experimental Procedure

This section explains the variables considered in the experiment, the general instructions, the output files obtained, and the post-questionnaire.

4.3.1. Experiment Variables

This section explains the different variables used in the experiment; independent, dependent and confounding.

4.3.1.1. *Independent Variables*

Independent variables are the variables controlled by the experimenter, those variables included:

- Experiment order: in order to eliminate fatigue and learning effects from the experiment three different orders were run as follows:

  Continuous-Discrete-Mixed (CDM)

  1) Discrete-Mixed-Continuous (DMC)
  2) Mixed-Continuous-Discrete (MCD)

The order was alternated for every participant. The fatigue effect could develop on disintegration of skilled performance, which leads to a loss of overall management control of the next scenario. In order to prevent this fatigue effect from being reflected in the final scores, every participant alternated his/her order to perform the experiment.

- Type of system (scenario): there were three different scenarios that were tested in the experiment (Continuous, Discrete and Mixed).

4.3.1.2. *Dependent Variables*

Dependent Variables are the variables measured in an experiment. These were:

- Subsystem attended to at any time: indicates which subsystem was operated in that time period (one tenth of a second).
- Weighted score
    - Total Weighted Average Score is the final score presented to the participant and is the calculation of the mean of cumulative scores of every subsystem .

> The objective of the TME operator is keep each subsystem status level in its satisfactory or green ($50\% \leq s \leq 100\%$) range, and not let it drop into the unsatisfactory or yellow ($10\% \leq s < 50\%$) range, or the very unsatisfactory or red ($0\% \leq x < 10\%$) range. The TME computes a CTM performance measure based on this objective. The instantaneous score for a subsystem at any time is **qi**,. The variable **q** is a qualitative transform of the subsystem's current status level; **q** = +1 if the subsystem's status is satisfactory, **q** = 0 if its status is unsatisfactory, and **q** = -1 if its status is very unsatisfactory. The variable **i** is the subsystem's importance, the number appearing directly below the subsystem's status bar in the interface. The cumulative score for the subsystem is the mean instantaneous score since the beginning of the run. The total weighted score is the summation of all subsystem cumulative scores and reflects an overall task management performance measure, weighted according to subsystem importance. (Nicolalde, Funk, Uttl, 2003)

- Percentage (%) of total maximum weighted score is the transformation of weighted score into a percentage value; i.e., a percentage of the maximum.

  - The score of each subsystem at any given time is represented by the height of the status bar in the raw data sheet.

- Shedding indicates the number of subsystems that the participant chose to not pay attention to and let fall to 0%
- Number of transitions indicates the total number of switches between subsystems.

4.3.1.3. Confounding Variables

Confounding variables are factors that modify the original conditions and are out of control in the experiment such as distractions, heath conditions, etc. Confounding variables could change the final results or scores, modifying the initial conditions in the experiment. Those variables included:

1. Computer skills: This is the knowledge or experience that a participant has using a computer or any kind of software.
2. Timing of the experiment: the experiments were conducted at different times of the day from 9 am until 6 pm. This could influence the awareness of the participants and the exhaustion level.
3. Date of the experiment: the experiments were run throughout the term. Thus, it is possible that some participants could have been stressed out by specific circumstances, for example, midterms or final exams.
4. Degree of stress of participants after class, before class, presentations, dates, tournament, interviews, etc. The participants could be relaxed or not.
5. Weather conditions: This could influence the behavior and awareness of the participants.

4.3.2. Pilot Study

The pilot study used six participants. This study was used to test if the three scenarios were approximately equivalent in difficulty level. Also, the pilot study checked the learning rate of participants to assure that participants reached the top of the learning curve by the fifth trial, which was the recording data trial. The results showed that the difficulty level was not equivalent among the three scenarios. Therefore, it was necessary to adjust the weight, DR and CR parameters in the three scenarios in order balance the degree of difficulty.

4.3.3. General Procedure

Participants were tested one by one in a single session lasting two hours maximum. The test was given in a computer lab located at the Industrial and Manufacturing Engineering Department at Oregon State University. The participants first read and signed the Informed Consent Document (Appendix 2). If the participant had any questions, the experimenter answered those immediately. Then, the experimenter gave a short presentation about multitasking and some background on the Task Management Environment (TME). The information on multitasking included examples of multitasking on driving or flying. Background of TME included some differences between systems (continuous, discrete and mixed) and examples of continuous and discrete tasks.

According to the order in which the participants participated in the experiment, the general instructions were given (Appendix 3) and the participant performed four practice trials of 5 minutes each in order to become familiar with the TME software. At the end of each trial, the participant's percentage (%) of Total Weighted Average Score was recorded in an Excel data sheet. The full data set was not saved. The program was reset and a new trial started. The fifth trial was saved. After the fifth trial, the participant took a five-minute period to rest.

After the rest period, the participants continued with the next set of trials with the three TME system scenarios. At the end of the third set of trials, the participant was asked to answer the post-experiment questionnaire (Appendix 4). The answers were recorded in an electronic data sheet. As was stated before, the experiment was run in three different ways. The sequence of continuous/discrete/mixed scenarios was changed once every three participants.

4.3.4. General Instructions

Depending on which order the experiment was run, the participant was given the following verbal instructions: First, the participants received the basic instructions and was given the goal of the experiment. Then, a short description of the scenario and how the scoring system works were explained (Numerals 1 and 4 in the instructions. See Appendix 3).

After the general instructions were given, participants running the continuous system scenario first received instructions about numeral 6 (how continuous subsystems behave). After running the first set, the participant received instructions from numeral 7 (discrete system behavior) and in the final mixed system received instructions from numeral 8 (about the arrangement or location of subsystems in the mixed scenario).

If participants were running the discrete system first, they received instructions about discrete subsystem behavior (numeral 7). Subsequently, they received instructions regarding continuous subsystem behavior and mixed scenario arrangements (numerals 6 and 8) in order to teach the participants about handling continuous and discrete subsystems.

Finally, if the participants were running the third test order with the mixed scenario first, they received instructions regarding how continuous and discrete subsystems behave, as well as, how mixed scenario arrangements work (numerals 6, 7 and 8.) This was done because the participants needed to know how to handle continuous and discrete subsystems at that time. However, in the second and third set, it was not necessary to give extra instructions since the participants already understood how continuous and discrete subsystems worked. Appendix 3 shows a sample of the instructions given to the participants.

4.3.5. Output Files

Three kinds of output files were automatically generated when the experimenter saved the data from a scenario run. The first file was a text data file (name.txt), which contained information about the scenario, types of subsystems, CRs, DRs, Weights, Partial Scores (score for each subsystem) and Weighted Score for all subsystems. This file also recorded the time span (in seconds), recorded the Total Weighted Average Score, and showed the behavior of the raw data of all the subsystems (every tenth of a second) and showed which subsystem was attended to. The second file was name tm1 and contained summary information from the text file, which included the subject ID, time and date record of the data, subsystem behavior, and weighted score. The last file was name tm2 and contained the raw data of all subsystems. All files were saved in text format, which allowed the use of any text processing program to work with them.

4.3.6. Post-Experiment Questionnaire

Once the participants finished the three scenarios, they answered a questionnaire about their experience working with computers, the TME program and some questions regarding strategies used for the different scenarios. The following questions are a sample from the questionnaire included in Appendix 4.

1 What strategy did you use for continuous/discrete/mix system (scenario)? (Pattern of Attention)

2 When you held the button, for how long did you hold it down? (For Continuous)

Until Top (100%)   Until Satisfactory Level (green Zone)   Other

4.4 Experimental Design

4.4.1. Learning Curve

The calculations of mean performance for each trial were plotted in order to test that a learning process had been completed and performance had stabilized. The ideal curve should have reached the top of the learning curve in the fifth trial (recording data trial) in an asymptotic mode. In order to prove that the last two trials reached a stable learning state, a t-test was performed. If the difference between the last two trials was not statistically significant, it could be assumed that the learning process reached the top of the learning curve and no further learning process took place.

4.4.2. Correlation Study

The main objective of this experiment was to determine a relationship or possible correlation between CTM performances on continuous, discrete, or mixed scenarios. The correlation study was performed based on the Percentage of Total Weighted Score. The first step was to collect the final scores from the raw data. The experimental design that was used was a crossover design. The chosen design is based on the hypothesis of mean score difference between two scenarios.

The second step was to formulate the experimental hypotheses. The hypotheses that were formulated were:

Null Hypothesis:	Corr(Continuous,Discrete) = 0
Alternative:	Corr(Continuous,Discrete) > 0, or
	Corr(Continuous,Discrete) < 0

Null Hypothesis:	Corr(Discrete,Mix) = 0
Alternative:	Corr(Discrete,Mix) > 0, or
	Corr(Discrete,Mix) < 0

Null Hypothesis: Corr(Continuous,Mix) = 0

Alternative: Corr(Continuous,Mix) > 0, or

Corr(Continuous,Mix) < 0

For the correlation study, the method used was multiple regression. This concept is closely related to a calculation of a correlation coefficient to measure the degree of association between two or more variables (Hayter, 1996.) When an experiment has two or more variables, it usual practice to find out how a particular variable (in this case is the scenario score) is dependent upon other variables. The analysis using this method results in the correlation coefficient (also known as the Pearson product moment correlation coefficient), which measures the degree of association between two variables and is denoted as r. Also the coefficient of determination is calculated, which is the square of the sample correlation coefficient; this variable is denoted as r^2. Finally, a residual analysis was performed in order to check if the regression was accurate and to check for evidence of homogeneity of variance and linearity.

The correlation study could be applied to all the data points or could be broken down by order groups (CDM-DMC-MCD). In order to decide what kind of analysis was necessary, the experimenter had to prove if there was a statistical difference between the order groups. Otherwise, the correlation study conducted was standard to all the participants' data scores.

4.4.3. Strategy Model

The secondary objective of this research was to determine whether or not there was any common strategy (an elaborate plan of action, according to the Oxford English Dictionary) or tactic (a planned action for accomplishing a goal) used by the participants in managing tasks. This study was limited to management of continuous tasks.

In order to describe this analysis there are some basic elements of strategy that need to be identified. These elements include task shedding, which shows the number of subsystems that the participant chose not to handle. Another element is the trajectory followed by participants in switching from one subsystem to another subsystem. One other element is the button down time that represents the time spent by a participant on each subsystem. Finally, the total time spent, which is the time holding down the mouse button in any subsystem.

1 Task Shedding is defined as how many tasks the participant allows to drop to zero. Figure 4.5 shows how a random participant shed subsystem S4 and let it drop to zero in order to only concentrate on subsystem S0. Graphics like this were used to determine the number of tasks that a participant shed in a run.

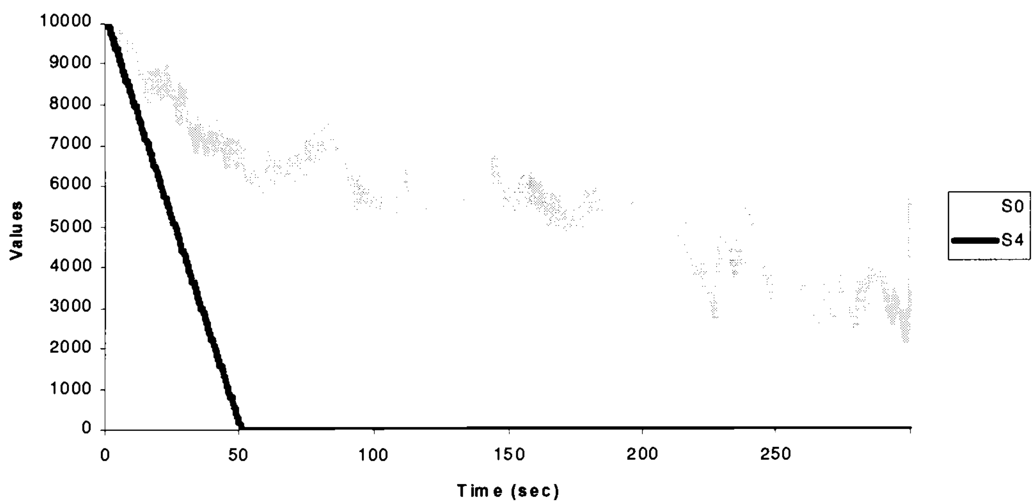

Figure 4.5 Shedding Tasks

2 Common path is the trajectory chosen by a participant to switch between subsystems throughout the experiment. Figure 4.6 shows a sample trajectory of the participants who had the best scores. These participants shed two tasks (less weight or important subsystems) and paid attention to the rest of the subsystems. The squares in the figure represent each of the subsystems

and the lines represent the trajectory that participant followed in switching among the subsystems.

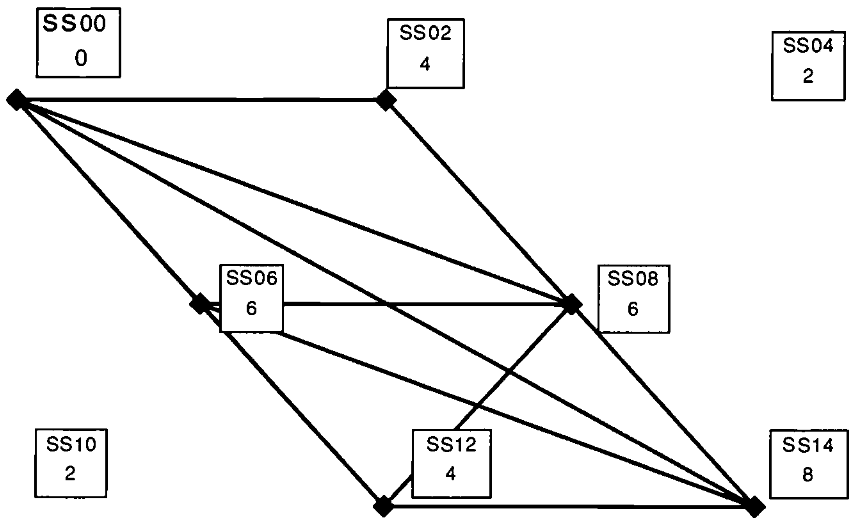

Figure 4.6 Common path top continuous scores

Button-Down Time was the time dedicated by participants to each of the subsystems. In other words, Button-Down Time is the total time that the participant held down the mouse button. Table 4.2 shows information regarding how many times a participant clicked the mouse in each subsystem, as well as the total time the participant held the button down (in seconds) and the percentage of time spent in each subsystem.

Table 4.2 Time spent per subsystem

Subsystem	Time Spent	Time Spent in sec	Time Spent in %
00	712	71.2	35.83
02	232	23.2	11.67
04	0	0	0
06	247	24.7	12.43
08	566	56.6	28.48
10	0	0	0
12	230	23	11.57
14	384	38.4	19.32
Sum	1987	198.7	100

3) Time distribution across subsystems is the proportion of time dedicated to each subsystem. Figure 4.7 is graphical representation of how much time a participant spent in a particular subsystem.

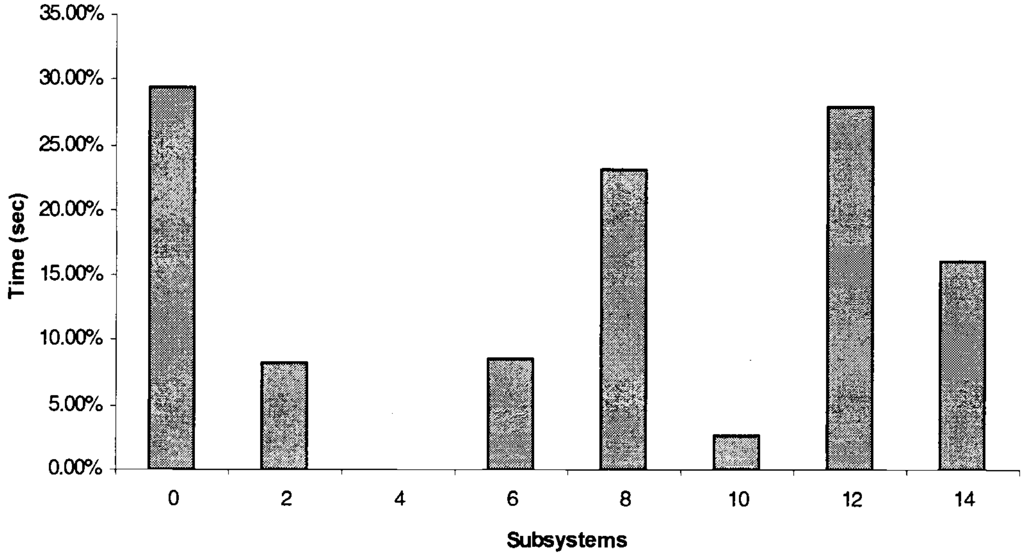

Figure 4.7 Time Spent on each subsystem

In Appendix 5 there are thoroughly explained procedures that describe how to find the number of tasks that participants shed, the button down time (time spent), and the most common path (trajectory plot.)

The above statistical analysis was performed to the Top (75%-100%) quartile and to the Bottom (0% - 25%) quartile of total weighted scores for continuous subsystem scenarios. A t-test of the difference between the mean scores of the top and bottom quartiles was performed.

CHAPTER 5 EXPERIMENTAL RESULTS

5.1. Overview

This chapter presents the results from the experiment described in the previous chapter. It contains learning curves for continuous, discrete, and mixed scenarios. Also, there is a comparison between the last training trial (number 4) and the recorded data trial (number 5.) Then descriptive statistics from final scores and top and bottom quartiles from continuous scores are presented. The correlation study between the three scenarios (continuous, discrete and mixed) is also presented. Finally, the strategy analysis for continuous scenarios is presented. This includes the percentiles division, the summary of data from top and bottom quartile, the time spent shedding tasks and transitions, and graphs and t-tests.

5.2. Results

5.2.1 Learning Curve Plots and Analysis

Figure 5.1. Shows the learning curve plot based on the mean results from the continuous scores. It can be observed that the last trial (recorded trial) reached the top of the learning curve. Figure 5.2 shows that the last practice trial (trial 4) and the record data trial (trial 5) have a difference that is statistically non-significant . The 95% confidence interval for the fourth trial is [68.4,72.7], and the 95% confidence interval for the recording trial is [70.4,74.2]. The t-test result was: t = -1.21 with a P-value = 0.227

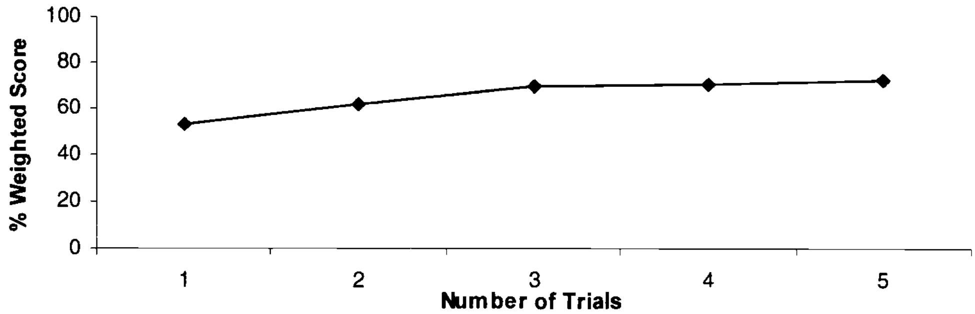

Figure 5.1 Learning curve plot for Continuous Scenario

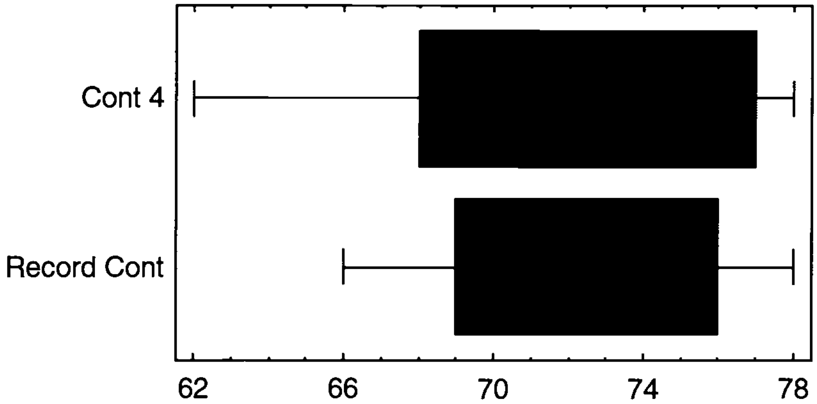

Figure 5.2 Comparison of last practice trial and record trial in continuous Scenario.

Figure 5.3 represents the learning curve of the scores from the discrete scenario. Figure 5.4 represents the comparison from the last practice trial and recorded data trial and shows that there is a difference that is statistically non-significant. The 95% confidence interval for the fourth trial is [80.3,84.8], and 95% confidence interval for the recording trial is [82.2,86.6]. The t-test result was: $t = -1.191$ with a P-value = 0.235

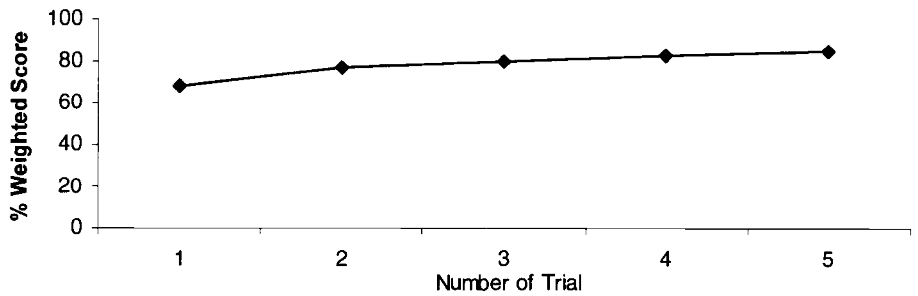

Figure 5.3 Learning Curve plot for Discrete Scenario

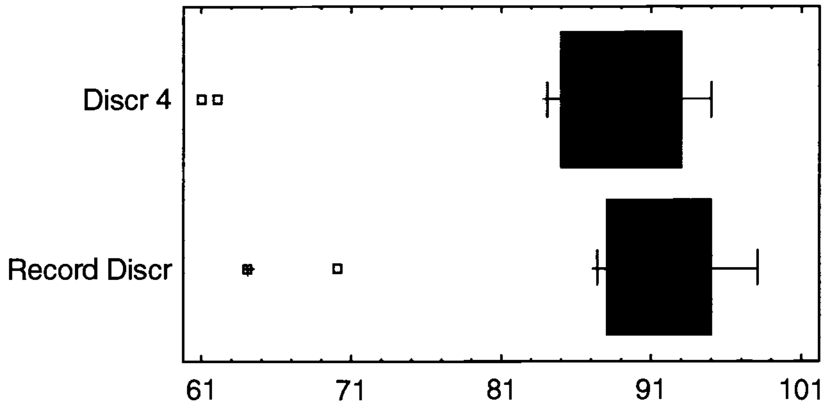

Figure 5.4 Comparison of last practice trial and record trial in Discrete Scenario.

Figure 5.5 represents the learning curve for the scores from the mixed scenario. Figure 5.6. illustrates the comparison from the last practice trial and the recorded data trial, which shows a statistically non-significant difference. The 95% confidence interval for the fourth trial is [72.1506,77.6627], and the 95% confidence interval for the recording trial is [74.4,79.4]. The t-test result was $t = -1.086$ with a P-value = 0.279

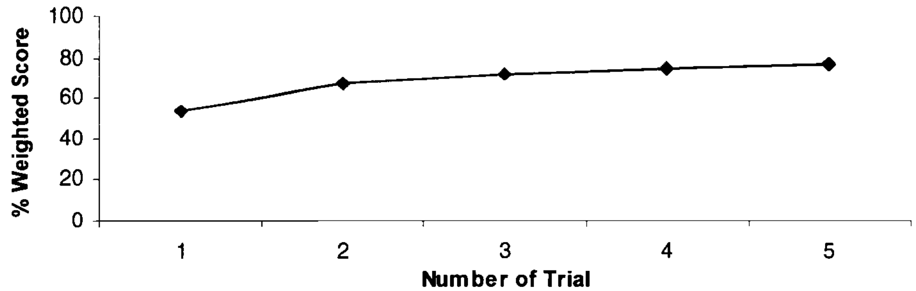

Figure 5.5 Learning Curve plot for Mixed Scenario

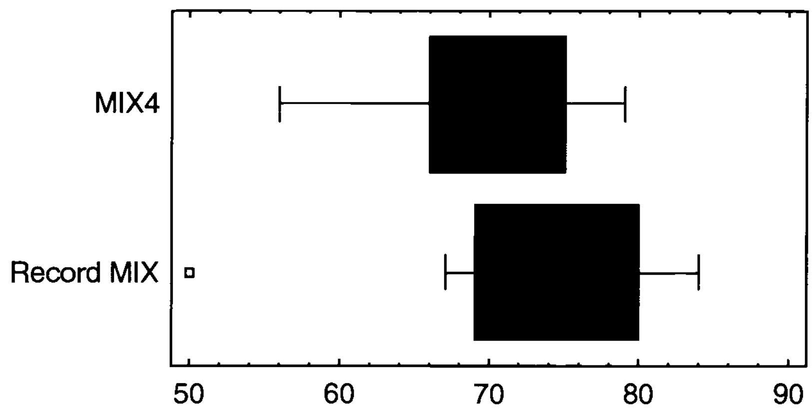

Figure 5.6 Comparison of last practice trial and record trial in Continuous Subsystem.

The previous learning curve studies demonstrated that the participant almost reached the top of the learning curve. However, there is still room for improvement, but it was not statistically significant. The data collected was suitable to perform a correlation study of final performance scores.

5.2.2 Descriptive Statistics

Table 5.1 represents the analysis of means, standard deviations, and maximum and minimum values for the final scores from all the TME scenarios.

Table 5.1 Descriptive Statistics for Percentage (%) Total Weighted Score

Scenario	Continuous	Discrete	Mixed
N	75	75	75
Mean	72.23	84.44	76.95
Stand D.	8.44	9.54	11.00
Max	86	98	93
Min	43	58	39

5.2.3 Correlation Study

Figure 5.7 shows the comparison of confidence intervals between order groups (CDM, DMC, MCD.) The comparison of means between the three groups shows that there was not a statistically significant difference between the three groups. The 95 % confidence interval for CDM was [231.8,243.4], the 95% confidence interval for DMC was [226.3,237.9] and the 95% confidence interval for the MCD was [225.4, 237.1]. The t-test result, was: t = 0.71 with a P-value = 0.4953. This indicates that the three order groups are not statistically different. Therefore, a correlation study could be conducted using all the participants' final results.

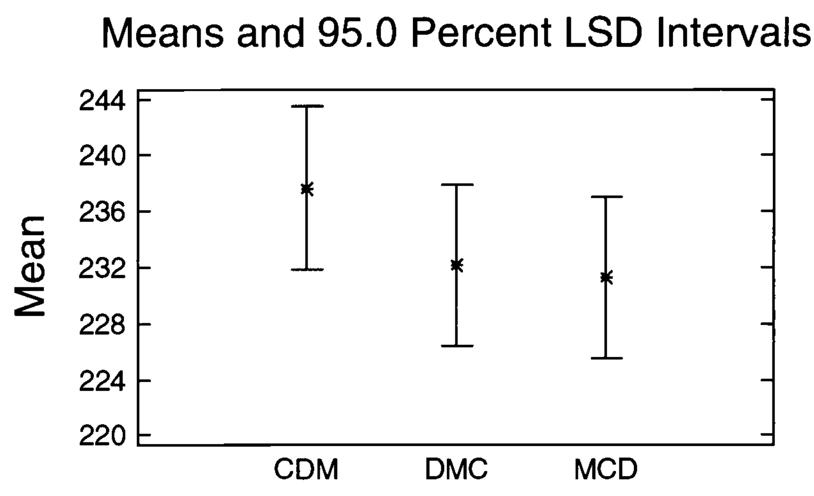

Figure 5.7 Means comparison for experiment order with a 95% confidence interval

Figure 5.8 shows the regression model of continuous scores vs. discrete scores. Since the P-value = 0.23 is greater than 0.05, there was not a statistically significant relationship between continuous and discrete scores at the 95% confidence level. The r^2 value explains that scores from continuous scenarios could predict scores from discrete scenarios up to 1.93 percent

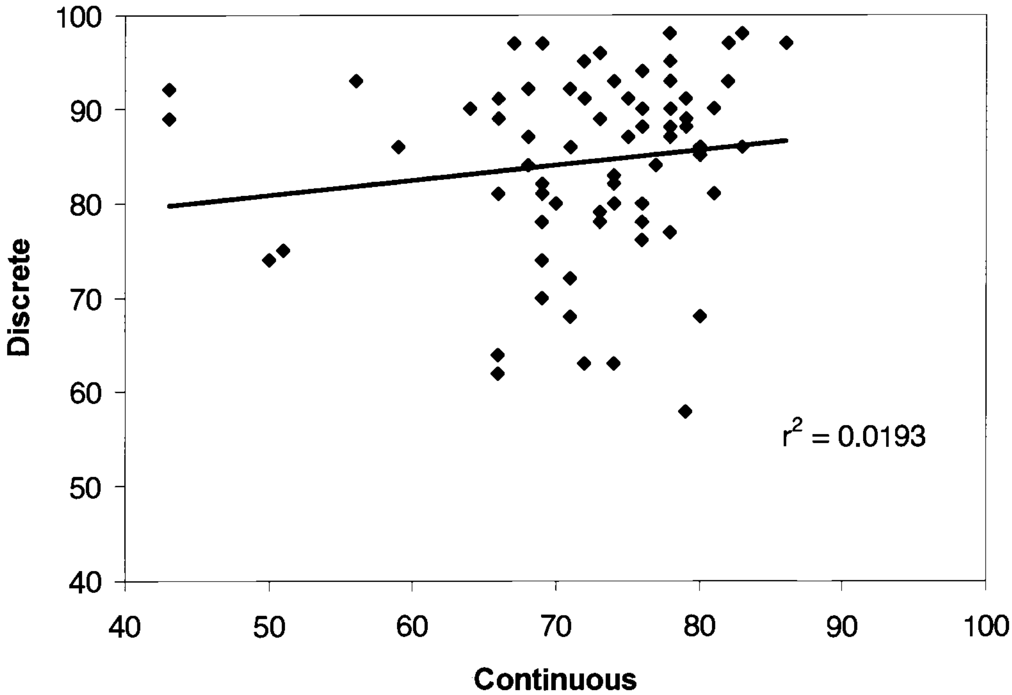

Figure 5.8 Regression model for continuous vs. discrete scores

Figure 5.9 shows the regression model of continuous scores vs. mixed scores. Since the p-value = 0.0087 is less than 0.05, this means that there is a statistically significant relationship between continuous scenario and mixed scenario scores at the 95% confidence level. The r^2 explains that scores from continuous scenarios could predict scores from mixed scenarios up to 9.1 percent.

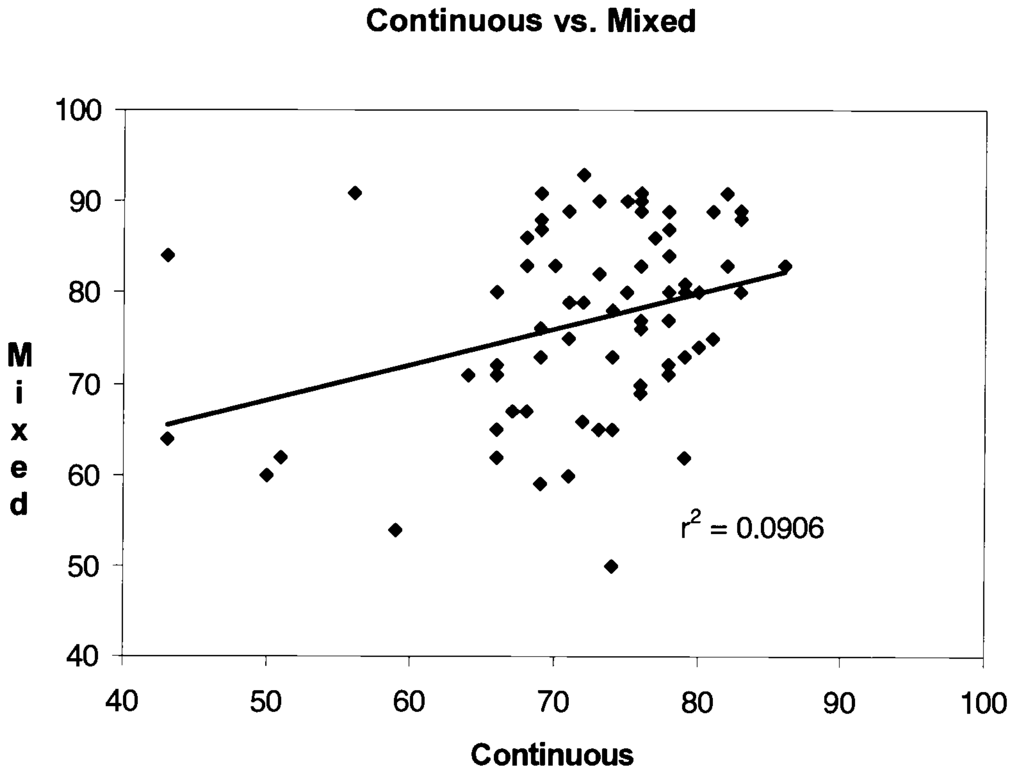

Figure 5.9 Regression model for continuous vs. mixed scores

Figure 5.10 shows the regression model of discrete scores vs. mixed scores. Since the P-value = 0.0133 is less than 0.05 it shows, that there is a statistically significant relationship between discrete and mixed scores at the 95% confidence level. The r^2 explains that scores from discrete scenarios could predict scores from mixed scenarios up to 8.1 percent

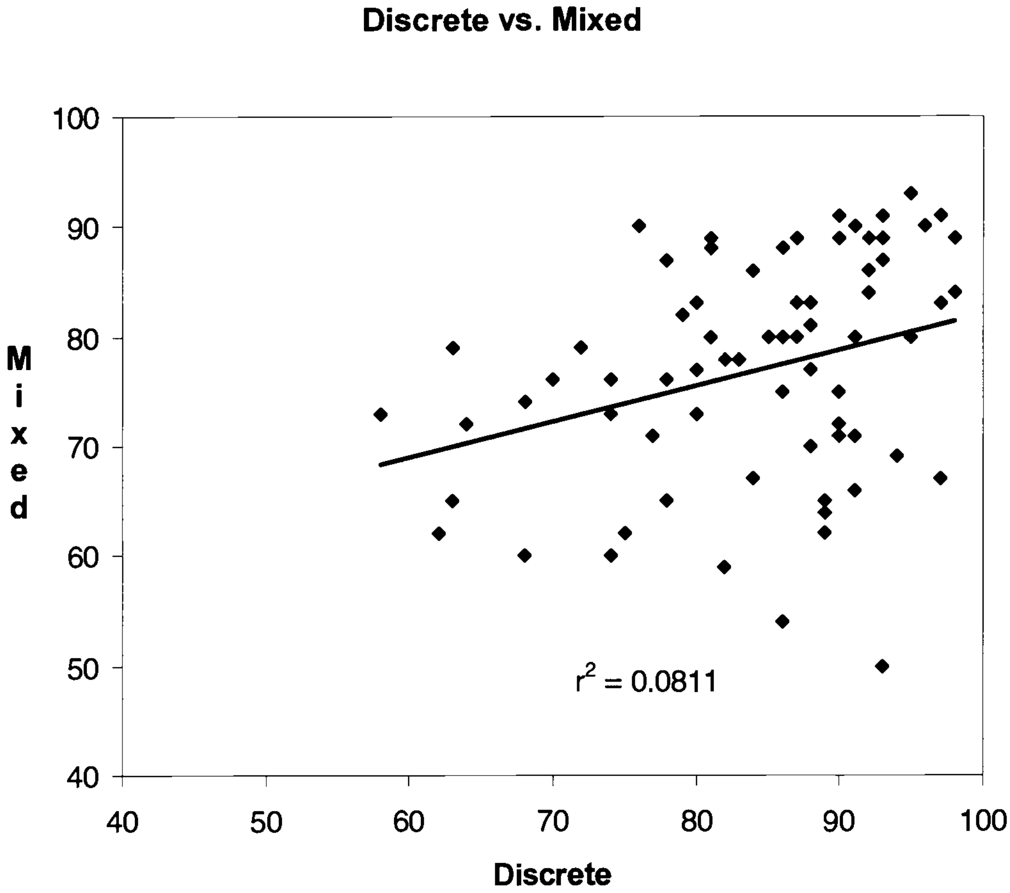

Figure 5.10 Regression model for discrete vs. mixed scores

Table 5.2 shows the correlation coefficients and the P-values of all scores from all participants. Based on the previous study of significant differences between group orders, this analysis is even more powerful since the sample is 75 participants.

Table 5.2 Correlation for General Data

	GENERAL		
	r	r^2	*P-value*
Corr. C vs. D	0.1389	0.0193	0.2346
Corr. C vs. M	0.3010	0.0906	0.0133
Corr. D vs. M	0.2848	0.0811	0.0870

Table 5.3 shows the results of a correlation analysis for the first group of participants that performed the experiment in the continuous-discrete-mix order. This analysis is less powerful than the general data analysis since the sample for this study is only 25 participants.

Table 5.3 Correlation for group continuous-discrete-mix

	CDM		
	r	r^2	*P-value*
Corr. C vs D	0.0718	0.0052	0.7332
Corr. C vs M	0.2801	0.0785	0.1722
Corr. D vs M	0.2819	0.0795	0.1750

Table 5.4 shows the results of a correlation analysis for the second group of participants that performed the experiment in the discrete-mix-continuous order. This analysis is less powerful than the general data analysis since the sample for this study is only 25 participants.

Table 5.4 Correlation for group discrete-mix-continuous

	DMC		
	r	r^2	*P-value*
Corr. C vs. D	0.2668	0.0712	0.1928
Corr. C vs. M	0.4335	0.1879	0.0304
Corr. D vs. M	0.4335	0.1198	0.0900

Table 5.5 shows the results of a correlation analysis for the third group of participants that performed the experiment in the mix-continuous-discrete order. This analysis is less powerful than to the general data analysis since the sample for this study is only 25 participants.

Table 5.5 Correlation for group mix-continuous-discrete

	MCD		
	r	r^2	*P-value*
Corr. C vs. D	0.1213	0.0147	0.5636
Corr. C vs. M	0.2211	0.0489	0.2881
Corr. D vs. M	0.2632	0.0693	0.2037

5.2.4 Strategy Study for Continuous Scenarios

Table 5.6 shows the data of the percentiles for continuous scores. From this table it is useful to extract the information for the top (100% - 75%) and bottom quartiles (0% - 25%.), which distinguishes the best and worst performers. The bottom quartile includes scores from 43 to 69, and the top quartile includes scores were from 78 to 86.

Table 5.6 Percentiles for Continuous Scores

1.0% --> 43.0
25.0% --> 69.0
50.0% -->74.0
75.0% -->78.0
99.0% -->86.0

Table 5.7 and Table 5.8 have the summary data information and the descriptive statistics for the top and the bottom quartile scores in the continuous scenario.

Table 5.7 Summary Continuous Data Top Quartile

ID	Score	Time Spent	Transitions	# of task shed
41	86	248.7	114	2
22	83	246.9	134	2
35	83	247.6	179	0
44	83	247	139	2
49	82	243.3	167	2
50	82	245.6	151	0
19	81	246.6	96	0
60	81	242.9	179	1
25	80	243.6	172	1
30	80	241.8	210	2
42	80	243.6	137	1
34	79	241.7	145	2
46	79	244.4	158	0
53	79	243.2	142	2
81	79	241.6	160	0
9	78	237	138	0
10	78	244.7	142	2
29	78	237.6	211	2
37	78	239.4	207	2
54	78	249	88	1
61	78	244.4	119	0
64	78	240.3	152	1
75	78	227.3	224	0
Mean	80.04	242.97	154.96	1.09
Stand. Dev.	2.23	4.70	35.65	0.90
Max	86	249	224	2
Min	78	227.3	88	0

Table 5.7 shows that participants at the top quartile had a mean of 242.97 sec out of a total of 300 sec of time spent using any subsystem (button down time), and in this instance, the participants made a mean of 154.96 transitions.

Table 5.8 Summary Continuous Data Bottom Quartile

ID	Score	Time Spent	Transitions	# of task shed
43	43	232.7	297	1
72	43	198.7	215	2
23	50	220.4	227	2
38	51	205.2	155	4
74	56	247.1	150	0
39	59	230.4	428	2
58	64	209.8	398	0
8	66	201.4	146	0
16	66	199.2	256	3
21	66	209.7	304	2
24	66	196.4	172	3
27	66	195.5	107	3
55	66	235.4	152	1
11	67	231.2	196	1
26	68	221.9	303	0
36	68	246.1	119	1
77	68	215.1	377	0
7	69	194.1	199	2
20	69	194.4	307	2
28	69	251.1	80	0
33	69	215.8	256	1
57	69	237.4	120	1
68	69	215	117	0
79	69	238.5	211	1
Mean	63.17	218.44	220.50	1.33
Stand. Dev.	8.28	18.44	96.80	1.17
Max	69	251.1	428	4
Min	43	194.1	80	0

Table 5.8 shows that participants at the bottom quartile had an opposite behavior from the top quartile participants. The mean was 218.44 sec out of 300 sec and a mean of 220.50 transitions.

Figure 5.11 shows that there is a statistically significant difference between the means of the bottom and top quartile variables at the 95% confidence level. The t-test result was: t = 0.34 with a P-value = 3.46E-9

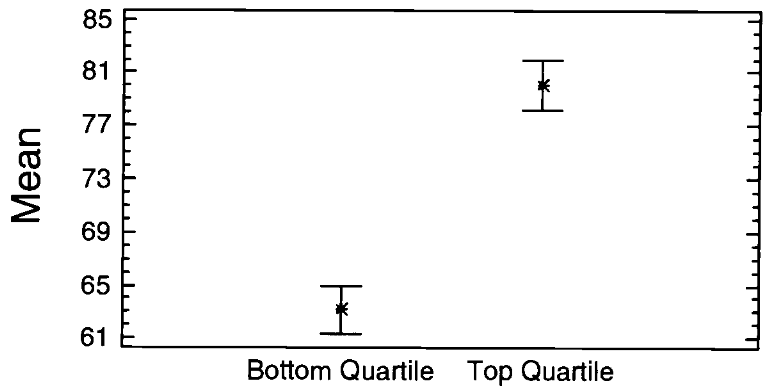

Figure 5.11 Means comparison for top and bottom quartiles in continuous scores

Figure 5.12 shows the time spent for the bottom group and top group. A means comparison study shows that top and bottom are different statistically. The 95 % confidence interval for the top is [240.9,244.9] and the 95% confidence interval for bottom is [210.6,226.2]. The t-test result is: t = -6.18 with a P-value = 4.96E-7.

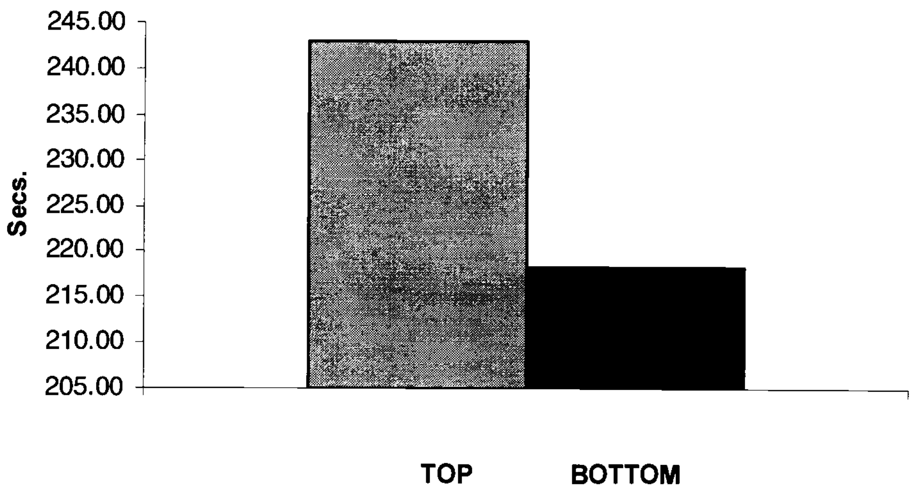

Figure 5.12 Total Time Spent

Figure 5.13 shows the mean graph of how many tasks are shed by the participants in the continuous scenario. The mean score comparison study shows that the top and bottom quartile were not different statistically. The 95% confidence interval for the top is [0.69,1.47] and the 95% confidence interval for bottom is [0.84,1.82]. The t-test result is: t = 0.81 with a P-value = 0.423

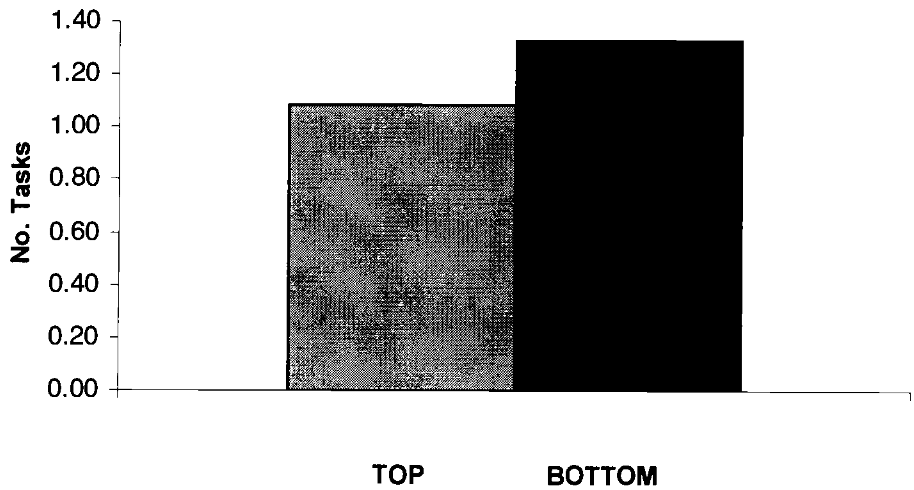

Figure 5.13 Shedding Tasks

Figure 5.14 shows the mean graph of how many transitions between tasks were made by the participants on a continuous scenario. The means comparison study showed that the top and bottom quartiles were statistically different. The 95% confidence interval for the top is [139.5,170.3] and the 95% confidence interval for bottom is [179.6,261.3]. The t-test result was: t = 3.05 and a P-value = 0.0037

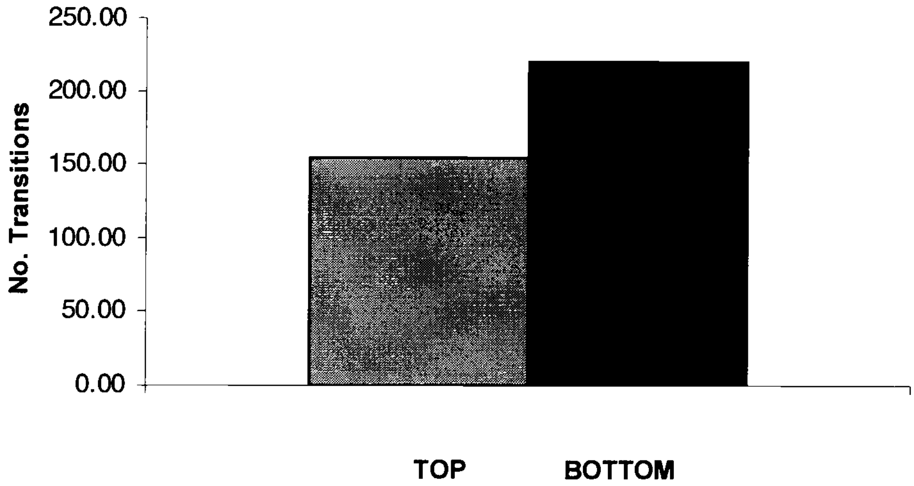

Figure 5.14 Total Transitions

The following figures show the difference between the paths follows by the best and worst performers in the continuous scenarios. The difference between figures 5.15 and 5.16 is the number of subsystems shed and also the number of transitions. According to the previous analysis of transitions, bottom quartile participants performed more transitions compared to the top quartile scores. These path comparisons were made graphically between the different sets of paths and by looking for common elements of trajectories of attention to subsystems. The main difference between the top quartile performers and the bottom quartile performers was the number of subsystems shed. The graphs also indicated that bottom quartiles had more transitions between subsystems, which could be calculated by the number of lines in each graph. The bottom quartile performers had two extra lines heading to subsystem SS10.

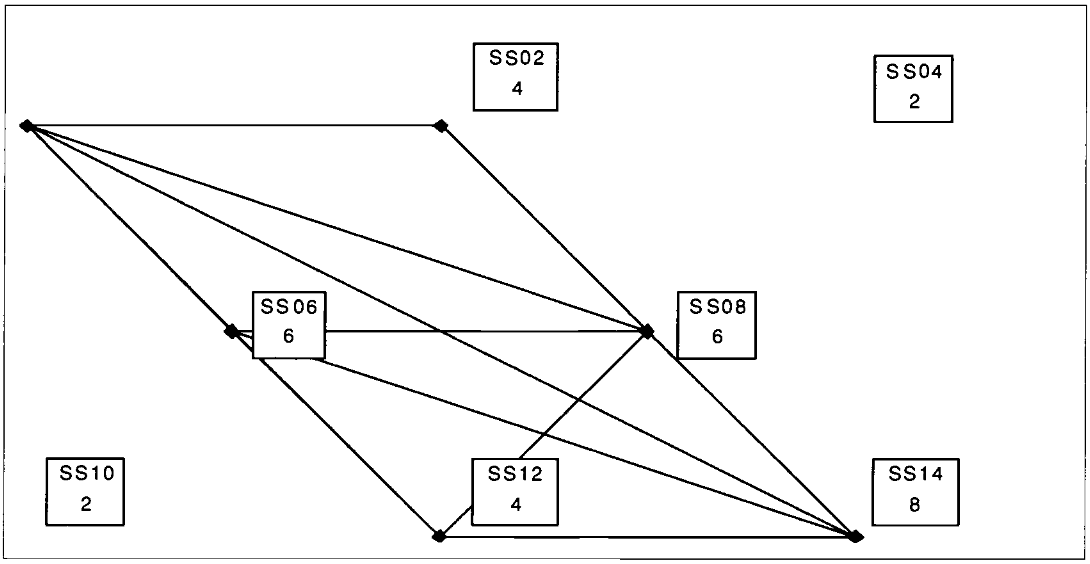

Figure 5.15. The path used by the top quartile performers.

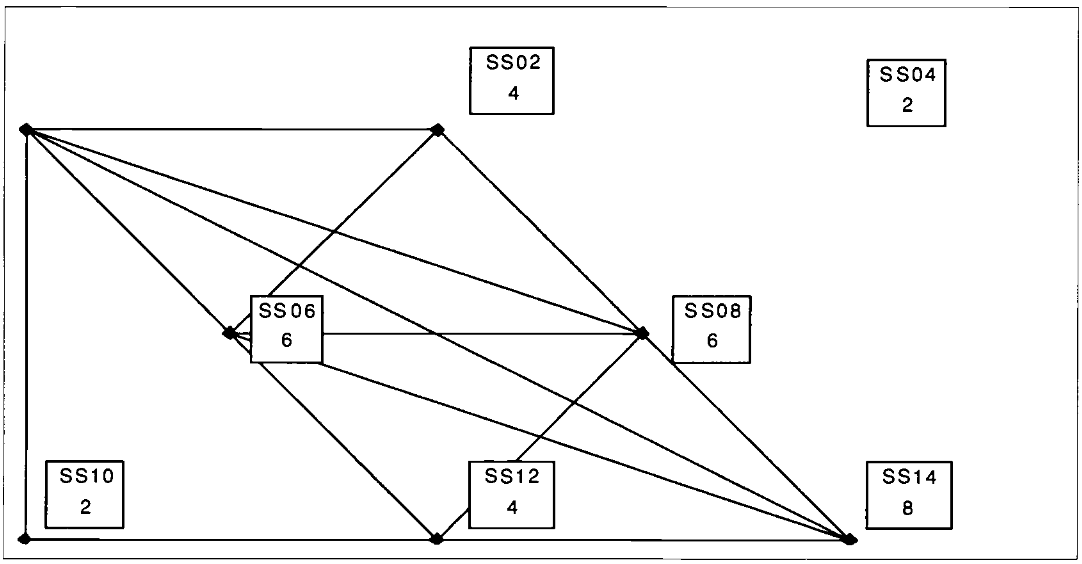

Figure 5.16. The path used by the bottom quartile performers.

CHAPTER 6 DISCUSSION

The purpose of this chapter is to interpret the research results presented in the previous chapter and to explain those results from the TME perspective. After that, there is an discussion of the implications of these results to the real world. Finally, the last section explains the limitations of this study.

6.1. Small Correlations between Scenarios

Why were there not significant correlations between the three scenarios, continuous, discrete, and mixed?

Each scenario required a different approach. Continuous subsystems required constant attention whereas discrete subsystems required a series of discrete actions. Therefore, with the discrete setup, attention was not constant as seen in the continuous subsystems. This implies that strategies required in one scenario were different than the ones require on another scenario. As a result, a participant could develop skills to handle one scenario correctly, but not necessarily other scenarios.

The scenarios' characteristics were very different from each other. Continuous subsystems were managed principally by the subsystem parameters of DR and CR. On the other hand, discrete subsystems are controlled by the parameters of DR and Discrete Maximum Count. Moreover, a mix of continuous subsystems and discrete subsystems implied a combination of parameters of CR, DR and Discrete Maximum account.

The result of the overall mean study showed that the subjects reached the top of the learning curve by the fifth trial (recording data trial). However, further investigation of group performance (for example the top and bottom quartile scores), revealed that there was a learning curve that reached the top and started to decline before the fifth trial (Figure 6.1). This could indicate an influence of fatigue in the performance of the bottom quartile scores. In addition, there were several comments from subjects in the post experiment questionnaire that mentioned fatigue or boredom due to the length of the experiment (two hours).

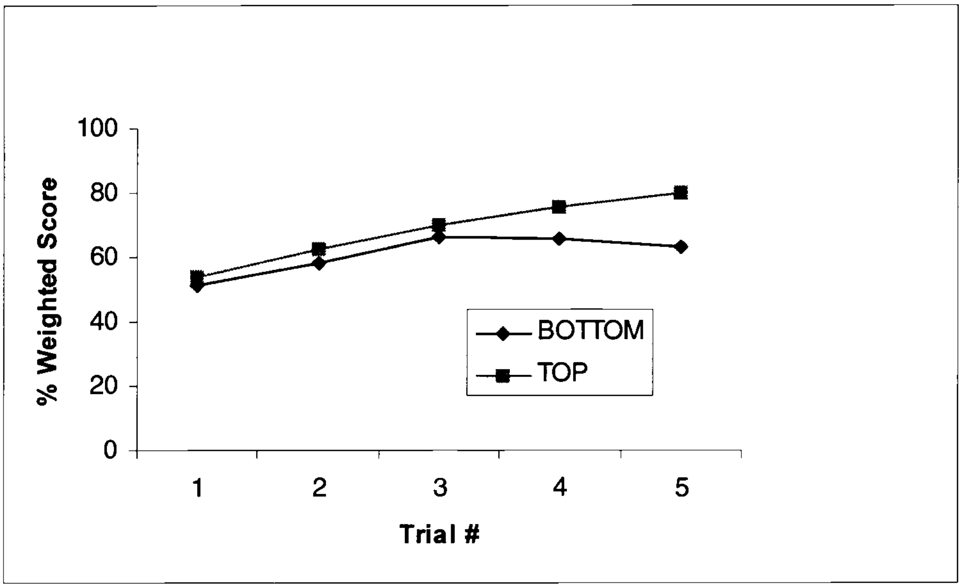

Figure 6.1 Learning curve for continuous scores (Bottom and Top Quartiles means)

There was a small significant correlation between Continuous-Mixed and Discrete-Mixed Scenarios. This is due to the fact that the mixed scenario was composed of 50% Continuous Subsystems and 50% Discrete Subsystems. These subsystems behaved in the same way as in the original purely continuous and purely discrete subsystems; the only difference was that there was a mix of 4 continuous and 4 discrete subsystems.

6.2. Discussion of Continuous Scenario Strategy

The analysis points up three remarkable findings: the differences in the number of subsystems shed between bottom and top scores, the difference in the number of transitions or switches between subsystems, and the differences in the time spent in each of the subsystems.

The number of subsystems that were shed by participants varied from one to four with an average of 1.09 for the top quartile and 1.33 for the bottom quartile. The most shed subsystem was SS 04 with a degree of importance (weight) of 2. This indicates that subjects correctly prioritized the tasks since they preferred to pay attention to the most important subsystems (with more weight) and shed the least important subsystems (SS02 SS04 SS10 SS12).

The top quartile subjects, which were the subjects with the best scores in the continuous scenario, seemed to focus less on transitions and to spend a longer time instead on each subsystem (more button down time.) On the other hand, the subjects with the worst scores represented in the bottom quartile, made more transitions and spent less time on each subsystem.

6.3. The Real World Implications

In order to have a good multitasking performance people have to prioritize and complete the most important tasks first. Additionally, it is necessary to focus on each task and spend the necessary time on it in order to avoid switching too frequently between different tasks. Switching tasks requires a time to abandon the current task being performed, and a time to review the status of a new task in order to resume working on it.

6.4. Scenario Suggested for the External Validation Study

Based on the results from the correlation study and the literature review, the most appropriate scenario to test for a validation study is a mixed scenario with continuous and discrete subsystems. The correlation study showed that there was a stronger correlation between either scenario and the mixed scenario. Also, the literature review showed that most concurrent task scenarios have a combination of continuous and discrete tasks. For example, flying an aircraft requires the pilot to perform multiple concurrent tasks, such as: Aviate task (Continuous), Navigate tasks (Continuous), Communicate Tasks (Discrete) and Manage System Tasks (Discrete and Continuous.) Another common example is driving a car, which implies performing different tasks including maneuvering the car (continuous), checking speed (discrete), paying attention to the road (continuous), operating the radio (discrete), and managing systems such as: lights, windshield wipers, AC or heater, (discrete.)

6.5. Limitations from the study

Although, the sample of the experiment was 75 participants, it was not necessary representative of the human population since the majority of the participants were engineering students attending Oregon State University. The sample was made of young collegiate adults between the ages of 19 and 22 years old.

There are several more analyses that could have been done on this data to look for interesting relationships among the different scenarios and subsystems, differences of performance by gender, career studies, stress level, etc. However, due to the time constraints , these analyses were not done.

6.6. Recommendations for Future Research

To reduce participant fatigue or boredom the experiment should be broken down into two sessions. The first session should be dedicated to training and getting familiar with the TME. The second session should be dedicated to recording data.

The length of the training trials and the recording data trials should remain the same. This is because the participants develop a strategy to manage the system for a certain amount of time. However, if the recording data trial has a different length of time, the participants have to change the strategy and this could influence the final performance.

In order to find out the optimal number of training trials necessary to reach the top of the learning curve, the researcher should run a pilot study to determine this number. In the case of this research, the learning peak was reached by the fourth trial. This means that the participants were ready to record data by the fifth trial.

REFERENCES

Aasman, J., Michon, J. (1992). Multitasking in driving. A cognitive architecture in perspective. Soar: A cognitive architecture in perspective: A tribute to Allen Newell. Studies in cognitive systems, Vol. 10. (pp. 169-198).

Colvin, K. (2000). Factors that affect task prioritization on the flight deck. Doctoral Dissertation. Oregon State University.

Crosby, J., Parkinson, S. (1979). A dual task investigation of pilots' skill level. Ergonomics Journal. Vol 22(12) Dec 1979, 1301-1313

Dismukes K., Loukopoulos, L., Jobe, K., (2001).The Challenges of Managing Concurrent and Deferred Tasks. International Symposium on Aviation Psychology, Columbus, Ohio. Dec 2001, 1-6

Funk K., Colvin K.,Bishara S, Nicolalde J, Shakeri S, Chen J., (2003).Training Pilots to Prioritize Tasks: Theoretical Foundations and Preliminary Experiments. Final Report.

Funk K., Braune R. (1999). The Agenda Manager: A Knowledge-Based System to Facilitate the Management of Flight Deck Activities. SAE International, and the American Institute of Aeronautics and Astronautics, Inc.

Hart, S., Hauser J., (1987). In flight application of three pilot workload measurement techniques. Aerospace and Environmental Medicine. Vol 58(5) May 1987, 402-410.

Hayer, Anthony J. (1996). Probability and Statistics for Engineers and Scientists. PWS Publishing Company.

Humbhrey D, Kramer, A. (1999). Age related differences in perceptual organization and selective attention. Dissertation Abstracts International.

Kantowitz, B.; Casper, P. A.(1987). Human workload in aviation. Cognitive Psychology. Academic Press series in cognition and perception. (pp. 157-187). San Diego, CA, US.

Lamoroux T. (1999). The Influence Of Aircraft Proximity Data On The Subjective Mental Workload Of Controllers In The Air Traffic Control Task. Ergonomics. Vol 42(11) Nov 1999, 1482-1491.

Loukopoulos, L., Dismukes, R., Barshi, I. (2003). Concurrent Task Demands in the cockpit: Challenges and vulnerabilities in routine flight operations. 12th International Symposium on aviation psychology. Dayton, OH. October 2003. 6p.

Milke, R. M.; Becker, J. T.; Lambrou, P.; Harris, H. C., Jr.; Schroeder, D J.. (1999).The Effects of Age and Practice on Aviation-Relevant Concurrent Task Performance. FAA Office of Aviation Medicine Reports. DOT-FAA-AM-99-22, Aug 1999, 16p.

Mongomery, Douglas C. (1997). Design and Analysis of Experiments. John Wiley & Sons, Inc.

Moray, Neville. (1979). Mental Work Load Its Theory and Measurement. Plenum Press.

Nicolalde, Roberto J., 2004. The Augmented Stage Model of Human Information Processing: How Well Do Cognitive Abilities Drawn from the Stages in the Model Predict Concurrent Task Management Performance. Masters Thesis, Oregon state University

Pasher, H. (1994). Dual-Task inference in simple tasks. Psychological Bulletin. Vol 116(2) Sep 1994, 220-244.

Prinzel, L., (2004). Head Up Displays and Attention Capture. NASA TM-2004-21300. February 2004.

Radeborg, K.; Briem, V.; Hedman, L. R. , (1999). The effect of concurrent task difficulty on working memory during simulated driving. Ergonomics Journal. Vol 42(5) May 1999, 767-777.

Rogers, W.H. (1996). Flight Deck Task Management: A cognitive Engineering Analysis. Proceedings of the Human Factors and Ergonomics Society 40th Annual Meeting. 4(3), 211-240.

Rubinstein J., Meyer D.,(2001). Is multitasking more efficient? Journal of Experimental Psychology: Human perception and performance. APA press releases. August 2001 Vol.27, No 4, 763-797.

Szalma, J.,Hancock, P (2003). Mental Resources and Performance Under Stress. Theoretical Issues in Ergonomics Science. Volume 4, Number 1-2 January-April 2003. 238 – 249.

Sodhi, M., Reimer, B., Llamazares, I. (2002). Glance analysis of driver movements to evaluated distraction. Behavior Research Methods, Instruments & Computers. Psychonomic Society. Vol 34(4) Nov 2002, 529-538.

Shakeri, S., 2002, A mathematical Modelling Framework for Scheduling and Managing Multiple Concurrent tasks, Doctoral dissertation, Oregon State University

Tsang, P., (1986). Can pilots time-share better than non-pilots?. Applied Ergonomics. Vol 17(4) Dec 1986, 284-290.

Ward, D. (1996). The measurement of driver's metal workload, Doctoral Dissertation. University of Groningen.

Wickens,C.,Raby, M., (1994). Strategic Workload Management and decision biases in aviation. The international journal of aviation psychology. Vol 4(3) 1994, 211-240.

Wickens,C., Gordon S., Liu Y., (1997). An Introduction to Human factors Engineering. Addison Wesley Longman, Inc. New York: Longman.

Wickens, C., Hollands J., (2000). Engineering psychology and Human Performance. Prentice Hall.

Wickens, C. (2002). Multiple Resources and performance prediction. Theorical Issues in Ergonomics Science. Vol 3(2) Apr-Jun 2002, 159-177.

Wiener, E., Nagel D., (1988). Human factors in Aviation. Academic Press Inc. (1988). xxi, 684 pp.

APPENDICES

Appendix 1 - Table Final Scores Percentage % of total weighted score

Subject ID	Continuous	Discrete	Mixed	Order
7	69	70	76	CDM
8	66	64	72	DMC
9	78	98	84	MCD
10	78	95	80	CDM
11	67	97	67	DMC
12	76	94	69	MCD
13	74	93	50	CDM
14	76	88	70	DMC
15	76	88	83	MCD
16	66	81	80	CDM
17	71	86	75	DMC
18	71	72	79	MCD
19	81	81	89	CDM
20	69	78	87	DMC
21	66	89	65	MCD
22	83	98	89	CDM
23	50	74	60	DMC
24	66	81	80	MCD
25	80	86	80	CDM
26	68	84	67	DMC
27	66	62	62	MCD
28	69	97	91	CDM
29	78	77	71	DMC
30	80	68	74	MCD
31	71	68	60	CDM
32	72	91	66	DMC
33	69	74	73	MCD
34	79	88	81	CDM
35	83	86	80	DMC
36	68	92	86	MCD
37	78	88	77	CDM
38	51	75	62	DMC
39	59	86	54	MCD
40	73	96	90	DMC
41	86	97	83	MCD
42	80	85	80	CDM
43	43	92	84	DMC
44	83	86	88	MCD
45	76	78	76	CDM
46	79	91	80	DMC
47	74	80	73	MCD
48	72	95	93	CDM
49	82	97	83	DMC
50	82	93	91	MCD
51	73	79	82	CDM
52	75	87	80	DMC
53	79	89	62	MCD

54	78	90	72	CDM
55	61	87	69	DMC
56	73	78	65	MCD
57	69	81	88	CDM
58	64	90	71	DMC
59	76	76	90	MCD
60	81	90	75	CDM
61	78	87	89	DMC
62	74	63	65	MCD
63	76	90	89	CDM
64	78	93	89	DMC
65	71	92	89	MCD
66	72	63	79	CDM
67	76	90	91	DMC
68	69	82	59	MCD
69	75	91	90	CDM
70	76	80	77	DMC
71	73	89	39	MCD
72	43	89	64	CDM
73	74	83	78	DMC
74	56	93	91	MCD
75	78	93	87	CDM
76	70	80	83	DMC
77	68	87	83	MCD
78	74	82	78	CDM
79	69	74	76	DMC
80	77	84	86	MCD
81	79	58	73	CDM

Appendix – 2 Consent Form Document

Oregon State University
Department of Industrial and Manufacturing Engineering

I hereby give my consent to participate in an experiment conducted by Carlos Vasquez under the supervision of Dr. Ken Funk of the Oregon State University Department of Industrial and Manufacturing Engineering.

Purpose. This project is being conducted to learn how people manage multiple, concurrent tasks, as when driving a car or flight an airplane.

Procedure. I understand that the experiment will be conducted as follows. I will receive training and practice for each of the three system scenarios and then I will operate the TME for data collection purposes. Each trial of practice and data collection takes around 30 minutes. After running the trials, I will fill out a questionnaire. No information identifying me will be recorded. The total length of the experiment should not be for more than two hours.

Risks I understand that the probability and magnitude of harm, inconvenience, or discomfort anticipated in this study are no greater than those encountered in daily life.

Benefits. I understand that I will receive no direct benefits from my participation in this experiment.

Compensation. I understand that I will not be paid for my participation in this experiment.

Voluntary Participation. I understand that my participation in this study is voluntary and that I can withdraw from the experiment at any time without any kind of penalty.

Confidentiality. I understand that the data collected in this study will be available to the research investigators, support staff, and any duly authorized research review committee. The data will be kept confidential to the extent permitted by law. The data will be kept for 5-6 years following the publication of the results (the usual time required for keeping original data and records).

I grant Oregon State University permission to reproduce and publish all records, notes, or data resulting from participation, provided there will **be no association of my name** with the collected data and that confidentiality is maintained unless specifically waived by me.

Questions. I understand that I will have the opportunity to ask questions and receive satisfactory answers from the graduate students conducting the experiment. I understand that any further questions concerning this experiment should be directed to Dr. Ken Funk at (541) 737-2357, funkk@engr.orst.edu.

If I have questions about my rights as a research participant, I should contact the Oregon State University Institutional Review Board (IRB) Human Protections Administrator, OSU Research Office, (541) 737-3437, IRB@oregonstate.edu.

My signature below indicates that I have read and that I understand the process described above and give my informed and voluntary consent to participate in this experiment. I understand that I will receive a copy of this consent form.

Signature of Participant ______________________________ Date: __________

Printed Name of Participant ____

Appendix 3 - General Instructions for the experiment

Your goal is to keep all bars in the satisfactory level represented by the green zone above the yellow zone of every bar.

There are three different systems scenarios (Discrete, continuous, and mixed). Each scenario is composed of eight subsystems distributed at equal distances between each other. Each subsystem is represented by: a vertical blue bar which height shows the status of that subsystem, the color zone bar, the subsystem importance number, the control button; and for discrete tasks a number that represents the number of cycles until recovery that is located under the control button.

3. You will be rewarded for keeping bars in the green zone and penalized for letting them drop into the yellow or red zones. These rewards and penalties are proportional to the subsystem importance numbers shown on the subsystem's control button.

4. To start the game you have to push the start button each time you ask to run the game. By the end of the run, you have to save the data choosing the yes option and record a code in the save data window.

5. In order to move to another subsystem, you will be able to hold down the mouse button while the cursor is being moved to another subsystem to keep the previous subsystem activated. Once you release the mouse button and click it again on the subsystem's control button where the cursor is located, now the new subsystem is activated.

6. Continuous subsystems behave as follows: when the program is running, the bars drop slowly at different rates set by the experimenter until you click the button near to the subsystem bar with the mouse. When the button is held down, the corresponding bar rises while the other bars continue dropping.

7. Discrete subsystems have a similar behavior to the continuous subsystems, with the exception that its status normally remains at 100%, even without operator attention until a random "failure" event occurs. At that time, the bar drops until the operator clicks the subsystem's control button. The decrease temporarily halts for a predetermined period of time, and the control button disappears temporarily. After that period, the control button reappears and the status keeps dropping until the operator clicks the control button and the decrease pauses again. This cycle continues for a predetermined number of cycles shown in a number located under the control button. When this number reaches zero (0), the status recovers to 100% and stays there until a new failure occurs.

8. Mixed system is composed by four continuous subsystems and four discrete subsystems. This system alternates subsystems starting with a continuous subsystem. The behavior of this system is similar to manage continuous and discrete subsystems.

Appendix 4- Post-Experiment Questionnaire

1. How long have you been using computers? ______ years

2. Have you used TME before?

   Yes No

3. How many subsystems did you plan to attend in average for all systems?

   1 2 3 4 5 6 7 8

4. What strategy do you use for continuous system? (Pattern of Attention)

5. What strategy do you use for discrete system? (Pattern of Attention)

6. What strategy do you use for mixed system? (Pattern of Attention)

7. When you hold the button, how long did you hold it down? (For Continuous)

   Until Top (100%) Until Satisfactory Level (green Zone) Other

8. Comments or suggestions you have concerning this experiment.

Appendix 5- Instructions How to obtain information from Excel Files

In order to find out the number of tasks that participants shed and the button-down time (Time Spent) it was necessary to run a macro program in Excel (TME-TAS-STRATEGY) that gives information about time spent in each subsystem and TIS Chart (Time spent in each Subsystem) that helps to visualize how much time was spent in each subsystem and how many subsystem the participant chooses to shed. In the first page of the Macro there is some information of how to input information to the Macro and how to run it.

The transition matrix was found through an Excel file (transition). This page has two worksheets pages that help to get the transition matrix for any file. The first step was to open any out text file from TME program. The experimenter has to cut the information columns that contain the Time (sec) and attended to. After this information has to paste in the transfer page and erase the blank spaces from attended to column. Following this process the information was copy from this page to the transition page and the transition matrix was process at the end of this page.

The most common path was originated from another Excel file (PLOT-PAGE). This page had a single page where information from any out put text file could be plot base on the paste of information from attended to column.