AN ABSTRACT OF THE
DISSERTATION OF

Amy Lynn Hoover for the degree of Doctor of Philosophy in Education presented on April 26, 2005.
Title: Experimental Analysis of Task Prioritization Training for a Group of University Flight Technology Students

Abstract Approved:

Redacted for Privacy

Darlene F. Russ-Eft

Task prioritization performance was evaluated for pilots who participated in a concurrent task management (CTM) training course and pilots who did not. CTM is the process by which pilots selectively attend to high priority tasks and shed non-priority tasks. Twenty seven pilots enrolled in a university flight technology program were randomly assigned to an experimental group and a control group. Pilots flew pretest and posttest simulated flights on an FAA approved flight training device (FTD). Twenty potential task prioritization errors were embedded at 14 locations within the flight scenarios. Pretest CTM performance of the two groups was comparable. During a two week period between pretest and posttest simulated flights pilots in the experimental group participated in a CTM training course designed and taught by an FAA certified flight instructor and pilots in the control group did not.

A Mann-Whitney U test rejected the null hypothesis that there was no difference in posttest CTM errors between the groups, indicating a positive training effect for experimental group pilots. Longer term training effects were not evaluated. Different cognitive processing models described various pilot behaviors; some behaviors were described by single channel theory, some by single resource theory,
and others by multiple resource theory. Mispriotization due to the interruption of an aviate task by a communicate task occurred more frequently than interruption of a navigate task by a communicate task. Fixation on the GPS navigational system caused more than half the pilots to deviate from primary aviate tasks to attend to the secondary navigate task.

Additional research with different participants is recommended. A study comparing training results between pilots who have different training backgrounds is also recommended. A longer time period between pretest and posttest and/or a longitudinal study is recommended to test for longer term training effects. Qualitative studies could also be used to enhance experiments, such as gathering responses from participants to discern the extent of their learning. Further studies using cockpits with higher levels of automation and complexity, such as new generation flat panel or 3D cockpit displays is recommended.
Experimental Analysis of Task Prioritization Training for a Group of University Flight Technology Students

by

Amy Lynn Hoover

A DISSERTATION

submitted to

Oregon State University

in partial fulfillment of
the requirements for the
degree of

Doctor of Philosophy

Presented April 26, 2005
Commencement June 2005
Doctor of Philosophy dissertation of Amy Lynn Hoover
Presented on April 26, 2005.

APPROVED:

Redacted for Privacy

Major Professor, representing Education

Redacted for Privacy

Dean of the School of Education

Redacted for Privacy

Dean of the Graduate School

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

Redacted for Privacy

Amy Lynn Hoover, Author
ACKNOWLEDGEMENTS

I am much indebted to my major professor, Dr. Darlene Russ-Eft, for her continued support and excellence in her critique of my work. Darlene’s input challenged me to bring greater depth and insight into my thinking and my writing. I also appreciate the unique contributions of each of the other committee members. Dr. Alex Sanchez gave sage advice early on that I referred back to often during the process. Dr. Kenneth Funk was an invaluable reference who provided the CTM theory, insight from previous studies, and numerous enlightening discussions. Dr. Jack Dugan was my statistics coach and moral support. Dr. Anthony Veltri assured that my process ran smoothly and provided thoughtful feedback in his critiques.

Special acknowledgement goes to students from the Central Washington University Flight Technology Program who participated in the study and to flight instructors Don McRade, Luke Reeves, Trevor Michael, Chris Sohren, and Teresa Sloan for their expert assistance. I greatly appreciate the continued support of my colleagues. Special thanks to Dale Wilson, Autumn Cooper, Teresa Sloan, William Bender, Barbara Masberg, Craig Johnson, and Beth Rogers.

A deep appreciation goes out to members of the OSU Community College Leadership Program Cohort 10. Brooke, Gweneth, Michelle, Ed, Lori, Andrea, Cam, Kal, and Angela: without you this would not have been possible and I thank you. My mother and father continuously supported and encouraged me throughout the process. I particularly appreciate my father’s perspective in helping me maintain my focus and my sense of humor. Finally, I want to thank Charles for taking me away when I needed to go and Des for always being there no matter what.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>1.1 Background</td>
<td>2</td>
</tr>
<tr>
<td>1.2 Previous Studies</td>
<td>8</td>
</tr>
<tr>
<td>1.3 This Study</td>
<td>9</td>
</tr>
<tr>
<td>1.4 Summary</td>
<td>11</td>
</tr>
<tr>
<td>2 LITERATURE REVIEW</td>
<td>12</td>
</tr>
<tr>
<td>2.1 Human Factors in Aviation</td>
<td>13</td>
</tr>
<tr>
<td>2.1.1 Historical Development</td>
<td>15</td>
</tr>
<tr>
<td>2.1.2 Human Error</td>
<td>18</td>
</tr>
<tr>
<td>2.1.3 Summary</td>
<td>19</td>
</tr>
<tr>
<td>2.2 Cognitive Theories of Multitasking Behavior</td>
<td>19</td>
</tr>
<tr>
<td>2.2.1 Single Channel Theory</td>
<td>20</td>
</tr>
<tr>
<td>2.2.2 Single Resource Theory</td>
<td>21</td>
</tr>
<tr>
<td>2.2.3 Multiple Resource Theory</td>
<td>22</td>
</tr>
<tr>
<td>2.2.4 Summary</td>
<td>23</td>
</tr>
<tr>
<td>2.3 Methods of Dealing with Cognitive Limitations</td>
<td>24</td>
</tr>
<tr>
<td>2.3.1 Attention Management</td>
<td>24</td>
</tr>
<tr>
<td>2.3.2 Workload Management</td>
<td>26</td>
</tr>
<tr>
<td>2.3.3 Task Management</td>
<td>27</td>
</tr>
<tr>
<td>2.3.4 Training in Attention, Workload, and Task Management</td>
<td>31</td>
</tr>
<tr>
<td>2.3.5 Summary</td>
<td>35</td>
</tr>
<tr>
<td>2.4 Concurrent Task Management (CTM)</td>
<td>35</td>
</tr>
<tr>
<td>2.4.1 Task Prioritization Factors</td>
<td>36</td>
</tr>
<tr>
<td>2.4.2 Technological Aiding of CTM in the Cockpit</td>
<td>41</td>
</tr>
</tbody>
</table>
TABLE OF CONTENTS (Continued)

2.4.3 Previous CTM Training Studies ............... 43
2.4.4 Summary ..................................................... 46

2.5 Learning Theories Relevant to Flight Instruction ...... 47

2.5.1 Learning Theory Foundations for CTM
       Training course ........................................... 51
2.5.2 Summary ..................................................... 56

2.6 Literature Review Summary ................................ 56

3 RESEARCH METHOD ................................................ 59

3.1 Hypotheses ........................................................ 59

3.2 Experimental Design .......................................... 60

3.3 Participants ....................................................... 62

3.4 Independent Variable ........................................... 68

3.5 Dependent Variable ............................................. 69

3.6 Internal Validity ................................................. 70

3.6.1 Pretest Variability and Differential Selection... 71
3.6.2 Practice Effect ............................................... 72
3.6.3 History Effect ............................................... 73
3.6.4 Experimenter Bias ........................................... 73
3.6.5 Mortality ....................................................... 75
3.6.6 Contamination ............................................... 76
3.6.7 Defining and Recording CTM Errors ............... 76
3.6.8 Internal Versus External Validity .................. 80
3.6.9 Summary ....................................................... 80

3.7 External Validity ................................................ 81

3.8 Experimental Platform ......................................... 83
TABLE OF CONTENTS (Continued)

3.9 Data Collection ................................................................. 85
3.10 Data Analysis ................................................................. 88
   3.10.1 Statistics .............................................................. 88
   3.10.2 Test for Homoscedasticity and Normal Distribution ... 90
3.11 Summary ................................................................. 92
4 RESULTS .............................................................................. 93
   4.1 CTM Error Data ............................................................ 93
   4.2 Null Hypothesis Test .................................................... 101
5 DISCUSSION AND CONCLUSIONS ...................................... 104
   5.1 Training Effect .......................................................... 104
   5.2 Multitasking Behaviors ............................................... 107
      5.2.1 Single Channel Theory ........................................ 107
      5.2.2 Single Resource Theory ...................................... 108
      5.2.3 Multiple Resource Theory ................................... 110
      5.2.4 Workload and Attention Management ............... 112
      5.2.5 Concurrent Task Management ......................... 113
   5.3 Conclusions ............................................................ 115
   5.4 Limitations .............................................................. 116
   5.5 Recommendations ................................................... 118
REFERENCES ............................................................................ 119
TABLE OF CONTENTS (Continued)

<table>
<thead>
<tr>
<th>APPENDICES</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Appendix A</td>
<td>Human Subjects Review Committee Approved Documents</td>
</tr>
<tr>
<td>Appendix B</td>
<td>CTM Training Course Syllabus and Outline</td>
</tr>
<tr>
<td>Appendix C</td>
<td>Line Oriented Flight Training (LOFT) profile Used for Pretest and Posttest FTD flights</td>
</tr>
<tr>
<td>Appendix D</td>
<td>Pilot Study</td>
</tr>
<tr>
<td>Figure</td>
<td>Description</td>
</tr>
<tr>
<td>--------</td>
<td>-----------------------------------------------------------------------------</td>
</tr>
<tr>
<td>2.1</td>
<td>Graph of task prioritization factors reprinted with permission from Funk et al. (2003)</td>
</tr>
<tr>
<td>2.2</td>
<td>Diagram showing the relationship of internal and external motivations to the flight student as social cognitive learner</td>
</tr>
<tr>
<td>3.1</td>
<td>Graph showing relationship of total flight time and pretest CTM error scores for all participants</td>
</tr>
<tr>
<td>3.2</td>
<td>Graph showing relationship of stage of training and pretest CTM error scores for all participants</td>
</tr>
<tr>
<td>3.3</td>
<td>Graph showing relationship of total instrument time and pretest CTM error scores for all participants</td>
</tr>
<tr>
<td>3.4</td>
<td>Graph showing relationship of total FTD time and pretest CTM error scores for all participants</td>
</tr>
<tr>
<td>3.5</td>
<td>Graph showing relationship of total Frasca 141 FTD time and pretest CTM error scores for all participants</td>
</tr>
<tr>
<td>3.6</td>
<td>Photo showing wide view of the Frasca 141 cockpit</td>
</tr>
<tr>
<td>3.7</td>
<td>Photo showing instructor station for the FRASCA 141</td>
</tr>
<tr>
<td>3.8</td>
<td>Histogram showing distribution of pretest-posttest difference in error scores for both groups</td>
</tr>
<tr>
<td>4.1</td>
<td>Graph showing the change in total CTM error scores for each group expressed as a percent of total possible errors</td>
</tr>
<tr>
<td>4.2</td>
<td>Graph showing pretest relationship of type of CTM error between control and experimental group</td>
</tr>
<tr>
<td>4.3</td>
<td>Graph showing posttest relationship of type of CTM error between control and experimental group</td>
</tr>
</tbody>
</table>
LIST OF FIGURES (Continued)

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.4.</td>
<td>Graph showing relationship of each type of error for the control group pretest and posttest flights</td>
<td>100</td>
</tr>
<tr>
<td>4.5.</td>
<td>Graph showing relationship of each type of error for the experimental group pretest and posttest flights</td>
<td>100</td>
</tr>
</tbody>
</table>
# LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1</td>
<td>Relationship of reported pretest criteria and pretest CTM error scores according to group</td>
<td>67</td>
</tr>
<tr>
<td>4.1</td>
<td>Frequency distribution of participant data and CTM error scores</td>
<td>94</td>
</tr>
<tr>
<td>4.2</td>
<td>Distribution of CTM error scores by type for the control group</td>
<td>96</td>
</tr>
<tr>
<td>4.3</td>
<td>Distribution of CTM error scores by type for the experimental group</td>
<td>97</td>
</tr>
<tr>
<td>4.4</td>
<td>Table showing raw scores and rankings used in the Mann-Whitney calculation</td>
<td>102</td>
</tr>
<tr>
<td>4.5</td>
<td>Results from SPSS calculation of Mann-Whitney test</td>
<td>103</td>
</tr>
</tbody>
</table>
1 INTRODUCTION

December 28, 1978 was a clear, calm night in Portland, Oregon, and certainly not the kind of tapestry against which one would think to paint the scene of a major airline accident. Unbeknownst to residents of the Rose City, the captain of a United Airlines DC-8 flying overhead became confused about his priorities. As the aircraft circled near the airport the captain became obsessed about a malfunction in the landing gear and allowed the aircraft to run out of fuel, even after other crew members warned him several times about the critically low fuel situation. The DC-8 crashed into a suburban neighborhood, destroying the aircraft and killing eight passengers and two crewmembers. The accident investigation report states that a major cause of the accident was the captain's "diverted attention from operation of aircraft" (NTSB DCA79AA005, 1979, p. 1). When the gear malfunction initially occurred it took priority. But as the airplane burned more and more fuel the situation changed and priorities shifted. When the fuel became critically low the captain neglected the most important task (ensuring the aircraft had sufficient fuel) to attend to the landing gear, a task he should have shed as it became less urgent with respect to immediate flight safety. That type of mistake constitutes a task prioritization error.

Like the United Airlines captain, all pilots are required to perform multiple tasks simultaneously during both normal and emergency operations. They must continuously assess, prioritize, execute, monitor, and shed tasks to the best of their ability, often in time critical situations and in a dynamic environment. This study focused on a critical question of whether pilots can change their task prioritization performance through training.

It is possible that human abilities to effectively process inputs and prioritize tasks vary greatly from one individual to another. However, a certain level of
multitasking ability is required in order to perform even at the basic level adequate to achieve initial pilot certification. As a result, pilots in general comprise a group that has been pre-selected to some extent based on their demonstrated ability to properly prioritize and execute tasks in the context of flight operations. Therefore, pilots may represent a more homogeneous group with respect to multitasking ability than a similar size group selected randomly from the general population. Pilots also undergo standardized training at each level of certification and rely heavily on standard operating procedures, use of checklists, and cockpit flow checks to help them prioritize tasks. Multitasking performance may indeed represent a combination of a pilot’s ability and training, and it seems reasonable that a system might be designed to help pilots better prioritize tasks in the cockpit. In this study a CTM training course was designed, delivered, and experimentally analyzed to assess whether or not pilots showed a change in their ability to effectively prioritize tasks in the cockpit after completing the training. Following is an overview of related background literature, previous studies, focus of this study, and an overview of the dissertation.

1.1 Background

Several theories of multitasking behavior are posited in the literature and explored in more depth in Chapter 2. Wickens (1992, 2002) defined multitask performance as the allocation of differentiated cognitive resources among competing tasks. Raby and Wickens (1994) explained multitask behavior as an attempt to manage workload and to balance acceptable levels of performance with acceptable levels of cognitive stress. A similar presentation of multitask performance as a function of workload was given by O'Hare and Roscoe (1990), but they related performance to the ability of the pilot to time share between concurrent tasks. Kern (1998) discussed effective execution of cockpit tasks as a function of proper procedural discipline in prioritizing both inputs and tasks.
Funk (1991) coined the term "Cockpit Task Management" in reference to a pilot's management of multiple concurrent tasks. He stated that pilots must manage concurrent tasks by assessing and prioritizing them, allocating resources in order of priority, and continuously updating their prioritization scheme. Funk, Colvin, Bishara, Nicolalde, Shakeri, Chen, and Braune (2003) described the cockpit as a multitask environment, where multiple, concurrent tasks compete for the pilot's attention and redefined the acronym CTM as "Concurrent Task Management,"(CTM). CTM is the process by which pilots selectively attend to tasks so as to complete the flight mission safely and effectively. Effective CTM is critical to flight safety and studies have found that CTM errors contributed to a significant number of aircraft incidents and accidents (Chou, Madhaven, & Funk, 1996; Damos, 1997; Dismukes, Loukopoulos, & Jobe, 1998; Latorella, 1996; Raby & Wickens, 1994; Rogers, 1996; Schutte & Trujillo, 1996).

Based on Funk's (1991) classification scheme, Chou, et al. (1996) conducted an in-depth review of 324 National Transportation Safety Board (NTSB) aircraft accident reports and 470 National Aeronautics and Space Administration (NASA) Aviation Safety Reporting System (ASRS) aircraft incident reports. They found that CTM errors were significant in 23% of the NTSB accidents and 49% of the ASRS reported incidents. Wilson and Funk (1998) conducted a study of commercial aircraft ASRS reports and noted that more technologically advanced aircraft showed a higher frequency of CTM errors. Funk et al. (2003) stated:

> From these studies, as well as the findings of Raby and Wickens (1994), Latorella (1996), Schutte and Trujillo (1996), Rogers (1996), Damos (1997), and Loukopolis et al. (2001, 2003) we conclude that CTM is a significant factor in flight safety. Experiments strongly suggest and adverse event studies confirm that CTM errors contribute to a significant number of aircraft incidents and accidents. (p. 6)

Wickens (2002) wrote, "Task management is directly related to mental workload as the competing demands of tasks for attention exceed the operator's limited resources" (p. 128). The inverse relationship of CTM performance to pilot workload was corroborated through empirical studies conducted by Chou, et al. (1996)
and those cited by O'Hare and Roscoe (1990) and Wickens, Dixon, and Chang (2003). Those studies found that pilots misprioritized tasks more frequently during periods of high workload, which demonstrates a type of CTM error called prioritization error. A prioritization error occurs when a pilot gives preferential attention to a lower priority task rather than to a task that should take higher priority with regards to flight safety (e.g., it is more critical, more urgent, or not being performed satisfactorily) (Funk, 1991).

One possible way to improve task management is through electronic aiding systems as recommended by Funk (1991) and Chou, et al., (1996). Two such systems were designed and tested by Funk et al. (2003) and are described in more detail in Chapter 2. However, they determined that technology required by such aiding systems was expensive, complex, not entirely reliable given current software limitations, difficult and costly to certify, and also added an additional component to the pilot's task load. They concluded that training pilots to better prioritize tasks was a possible alternative (Funk, et al., 2003).

Pilot multitasking is one aspect of what the Federal Aviation Administration (FAA) terms Aeronautical Decision-Making (ADM). ADM is defined as "a systematic approach to the mental process used by aircraft pilots to consistently determine the best course of action in response to a given set of circumstances" (Federal Aviation Administration, 1991, p. ii). The FAA acknowledged the importance of ADM in an advisory circular:

> Aviation has reached a new plateau. Acquiring aeronautical knowledge, airmanship skills, and proficiency are relatively easy. Navigation has been reduced to calculator simplicity. Modern autopilots and electronic displays have significantly reduced a pilot's workload. Today's technology requires administrative management and aeronautical decision making skills as prerequisites for safety and efficiency (FAA, 1991, p. 25).

Even more so than in 1991, pilots today operate in a multitask environment that becomes more complex as both aircraft and flight environment become more complicated. ADM comprises one element of the broader realm of study known as
aviation Human Factors. Human Factors focuses on explaining how the interface between humans and a variety of complex systems affect pilot and flight crew performance in "a time-constrained environment under stress" (Kern, 2001, p. 4).

Research and application of Human Factors in aviation focuses on the juxtaposition of pilots and the myriad of systems with which they continuously interact. Those complex systems include not only machines but also elements such as the flight environment and interactions with other individuals (i.e., weather, terrain, air traffic control, and internal and external communications). Additionally, psychological and psychosocial factors (i.e., workload, stress, situational awareness, cultural environment, and personal attitude) and physiological factors (i.e., illness, fatigue, fitness, diet, and discomfort) play a critical role in a pilot's ability to complete the mission safely and effectively (Federal Aviation Administration, 1991; Hawkins, 1993; Helmreich, 1998; Helmreich & Foushee, 1993; Kern, 1998, 2001; Reason, 1990; Shappell & Wiegmann, 2000).

Major world airlines and all U.S. airlines are now required to incorporate Human Factors and ADM into their training curricula (International Civil Aviation Organization, 1989; FAA, 1991). Helmreich and Wilhelm (1991) showed that those who have had such training outperformed their untrained counterparts. Additionally, U.S. military pilots are indoctrinated in elements of Human Factors beginning with their earliest training (Chappell, 1998; Kern, 1998, 2001; O'Hare & Roscoe, 1990). As a result of Human Factors training significant improvements in flight safety and a reduction in incidents and accidents have occurred in the military and airline industries in the past 20 years (Diehl, 2001; Helmreich, 1997; Helmreich & Foushee, 1993; Helmreich, Merritt, & Wilhelm, 1999; Kern, 2001). Research and evaluation worldwide shows that training commercial and military pilots and flight crews to detect, recognize, and manage errors originating from multiple regimes in a time-constrained environment is effective (Diehl, 2001; FAA, 2002; Helmreich, 1993; Helmreich & Foushee, 1993; Helmreich & Merritt, 2000; Kern, 1998, 2001; Klinek, Wilhelm, & Helmreich, 1999; Maurino, 1996; Maurino, Reason, Johnston, & Lee, 1995; Reason, 1990; Wiegmann & Shappell, 2001). The world's governments,
military, airlines, aerospace programs, and international aviation organizations have done extensive research and spent billions of dollars and in an ongoing effort to design, test, and utilize such programs.

In contrast, formal training, research, and evaluation for similar programs in civilian flight training conducted by the general aviation industry is almost non-existent (Diehl, 2001; Kern, 2001). General aviation comprises all civilian flying except scheduled passenger airlines and includes all non-airline civilian flight training. General aviation involves half a million pilots, includes more than 40 million flights per year, and includes over three quarters of all flying in the United States (Aircraft Owners and Pilots Association, 2004; Wells & Wensveen, 2004). The safety record of general aviation is much worse than that of the airlines. The National Transportation Safety Board reports a rate of 6.71 accidents per 100,000 flight hours for general aviation as compared with 0.3 per 100,000 for airlines (NTSB, 2004). Kern (2001) pointed out that in regards to general aviation accidents:

Although there has been slight improvement in number of fatalities per year, there has been no major statistical safety improvement since 1985. Although the trend is in the right direction, the pace of improvement needs a kick start. Because, when it comes to flying, general aviation is the most dangerous game in town. (p. 166)

Although the International Civil Aviation Organization recommended 15 years ago that all newly licensed pilots receive Human Factors training (ICAO, 1989) and the U.S. Federal Aviation Administration suggested training for pilots at all levels (FAA, 1991), the incorporation of such training into civilian flight schools curricula has just begun to catch on during the past half decade (Diehl, 2001; Jeppesen, 2003; Kern, 2001). One solution might be to use military Human Factors training for general aviation pilots. However, military training is highly specialized and involves special aircraft, systems, circumstances, and restrictions, and military training courses are not suited for training civilian pilots (Kern, 1998).

As a result of the need to improve general aviation flight training, the FAA has placed strong emphasis in the last decade on integration of ADM into flight training, from initial private pilot training through advanced operations. Research shows it is
critical for pilots to obtain proper training as early as possible to instill correct attitudes and behaviors that will carry through to all their flying (Buch & Diehl, 1983; Connolly & Blackwell, 1987; Diehl & Lester, 1987 quoted in Diehl, 2001). Therefore, flight schools and collegiate flight training departments are faced with the challenge of integrating ADM into their training course outlines (TCOs) to effectively prepare students to operate in a complex flight environment with state-of-the-art equipment. An effective means of assessing outcomes of CTM training as an important component of ADM could contribute to any general aviation flight training curriculum, including collegiate flight programs.

Before proceeding further it is important to clarify nomenclature encountered in the literature surrounding aviation Human Factors training. Training programs have evolved under different names at different airlines, military, and governing agencies. One commonly used term for training programs is Crew Resource Management (CRM). Kern (2001) wrote "CRM training is designed to produce team members who consistently use sound judgment, make quality decisions, and access all required resources, under stressful conditions in a time-constrained environment" (p. 4). As defined by Funk (1991), CTM is unique and distinct from CRM training:

CTM is distinct from Crew Resource Management (CRM). CRM has been defined as ‘using all available resources - information, equipment, and people - to achieve safe and efficient flight operations’ (Wiener and Nagel, 1988; Wiener, Kanki, and Helmreich, 1993). CRM includes optimizing interpersonal activities including leadership, effective team formation and maintenance, problem-solving, decision-making, and maintaining of overall situational awareness. CRM represents a focus on crew-level, as opposed to individual-level, aspects of training and operations.

In contrast to CRM, CTM places the emphasis on the individual’s training and performance. Based on conventional training and operational experience pilots are assumed to be able to do CTM. However, incident and accident data clearly indicate that this is not always the case, emphasizing the need for additional research and development. (Funk et al., 2003, p. 6)

Because this study involved the design and delivery of the CTM training course as a component of ADM, it is important to relate it to current progress in ADM
training and experimental assessment. Diehl (2001) wrote "Unlike CRM that was primarily employed for airline and military operations, ADM was initially applied to general aviation student training situations. The latter environment obviously involves relatively high error rates and low costs, thus permitting the use of controlled experiments" (p. 44). Diehl went on to explain that only six independent evaluations of ADM training worldwide were reported in the literature and that the success of such programs varied widely depending on the type of training conducted. He also noted that:

As expected, the effectiveness of the ADM materials varied widely depending primarily upon the comprehensiveness of the training. For the six studies, the improvement ranged from 8% in a voluntary, minimally structured situation to 46% for a well structured, comprehensive, ground school environment with simulator training. (Diehl, 2001, p. 44)

Thus any ADM training program, including CTM training, should be designed and implemented by qualified aviation educators and flight instructors and should include standard procedures and practices common to training programs that meet industry standards.

1.2 Previous Studies

Experimental results with respect to improving pilot CTM performance were somewhat inconclusive in two previous studies (Funk et. al., 2003). This investigator visited Funk et. al.'s (2003) experimental test site several times and had numerous discussions and written correspondences with Funk regarding those results. During those discussions, Funk acknowledged that the results could have been due to several factors, which are discussed in Chapters 2 and 3. Briefly, those factors included a dissimilar group of pilots who varied in experience from private pilots with low flight hours to highly experienced airline pilots with thousands of flight hours, and small sample groups (four subjects in each group).
A second limitation of previous studies was that the experiments were conducted using a popular PC based flight simulator game (Microsoft Flight Simulator) that is not approved by the FAA for any type of flight or procedures training. The components used to operate the simulator (for example a computer mouse) were also not of standard aircraft configuration and therefore unrealistic in simulating actual cockpit flight controls. Participants who had operated the computer game previously may have had an advantage, and other pilots might have exhibited improvement in task performance due to a practice effect as they became more familiar with Microsoft Flight Simulator.

The training materials used in Funk et. al.'s (2003) studies were not developed by a qualified aviation training specialist and did not follow standard practices and procedures as designated by the FAA and the industry. The training course was also not grounded in any established theory of teaching and learning and was not conducted by a qualified or certificated flight instructor. Finally, the specific CTM challenges defined by the experimental studies were sometimes inconsistent with standard procedures as defined by the aviation industry and the FAA (e.g., checklists, flow checks, practical test standards, and normal vs. emergency procedures).

1.3 This Study

This study analyzed the effectiveness of a CTM short course administered to a group of university flight students. The study was greatly aided by knowledge and experience from Funk et. al.’s (2003) studies and many of their recommendations were addressed in the design and implementation of this study.

Because training literature specific to CTM as defined by Funk et al. (2003) is scarce, the training course in this study relied heavily on other types of ADM and Human Factors training. Since a large body of training literature already exists in the general aviation industry this study took full advantage of those industry resources. Specifics of the training course are discussed in Chapter 2 and Appendix B.
The experimental design, methods, and results used in this study are discussed in Chapters 3 and 4, but a brief overview of those areas is provided here to clarify the significance and focus of the study and to guide the literature review discussion in Chapter 2.

To control for pretest variation in multitasking abilities between pilots, experiments comprised a pretest-posttest control group design with random assignment. To better control for differences in flight experience participants comprised a relatively homogeneous group of university flight students currently enrolled in a professional pilot degree program who were all undergoing the same course of training. Participants averaged 190 total flight hours, 56 hours instrument time, and 15 hours on the Frasca 141 flight training device (FTD) used for the study.

To enhance both internal and external validity, the CTM training course was grounded in appropriate learning theory and designed and delivered by a qualified and experienced flight instructor. Additionally, CTM was treated as an integral component of ADM and placed in perspective as a component of aviation Human Factors training. Design of the training course followed common practice and procedures for design of FAA approved training curricula. The course was taught using practices similar to those with which participants in the study were already accustomed. Desired training outcomes were articulated in the course training syllabus (Appendix B). Additionally, the FTDs used were FAA approved for logging instrument time toward pilot licensing and instrument certification.

Task prioritization was evaluated at specific locations, or challenge points, throughout the flight. Specific criteria for the challenge points was based on observations of actual errors that pilots made during the pilot study (Appendix D) and definition of performance errors followed published criteria from standard operating procedures and FAA practical test standards. All simulated flights as well as evaluation of errors were conducted by experienced FAA certificated instrument flight instructors (CFIIs).
1.4 Summary

Because the literature suggests that a significant number of aircraft accidents and incidents are attributed to CTM errors, this study was driven by the question of whether or not pilots would show improvement in task prioritization performance after completing a CTM training course. Because military and airline Human Factors training has been correlated to a reduction in aircraft accidents and incidents, a goal of this study was to use the analysis of the CTM training course to contribute to collective knowledge regarding effectiveness of ADM training and to inform the general aviation flight training community.

Previous experimental analyses of CTM training were ambiguous, but those programs were not designed or conducted by qualified aviation education professionals. Additionally, equipment used in previous studies (MS Flight Simulator) was not FAA approved, criteria for assessing the training did not follow standard FAA procedures and standards, sample sizes were small, and pilots had variable background and flight experience.

This study used recommendations from previous studies in conjunction with current knowledge from aviation Human Factors literature and ADM training to design and conduct the CTM training course. Additionally, the study used FAA and industry approved performance standards, practices, and FTDs to enhance experimental validity. The hope is that results from this analysis will contribute to knowledge critical to reducing errors and enhancing flight safety as well as posing questions for more in depth follow-on studies. The ultimate goal is to contribute to knowledge used for reducing error and enhancing flight safety.
2 LITERATURE REVIEW

The objective of this literature review is to critically analyze the historical and current theory, research, and writing relevant to the study. The review begins with a broad-brush account of the history and evolution of aviation Human Factors, including a brief look at human error related to flight operations in order to set the stage for more specific errors discussed later relative to ADM and CTM. Following that is a review of literature related to multitasking, including cognitive processes and methods of dealing with cognitive limitations, including attention management, task management, and training. Next is a review of literature related to CTM research, including factors that affect task prioritization and technological CTM training aids developed by researchers at Oregon State University, followed by a discussion of previous CTM training studies with respect to their techniques and outcomes. The final section of the literature review focuses on learning theories and practices relative to flight instruction and to the design and delivery of CTM training for the proposed study.

Research for the literature review took several forms. This investigator has been flight instructing, teaching aviation Human Factors, and conducting various ADM training courses for 13 years and is already familiar with many of the major resources in the field. Those resources were used to locate further references. Additionally, references cited in previous CTM studies, particularly Funk et. al. (2003), were researched. References pertinent to learning theories began with resources obtained through instructional leadership courses at Oregon State University and continued through investigation of secondary references, library searches, database searches, and recommendations from committee members.

Several computer database searches were conducted between December, 2002, and December, 2004 and included the following sites: Federal Aviation Administration (FAA), International Civil Aviation Authority (ICAO), National Aeronautics and Space Administration (NASA) Ames Research Center, The University of Texas at Austin Human Factors Research Project, The University of
Illinois at Urbana-Champaign Aviation Human Factors Division, The Australian Aviation Psychology Association, Academic Search Elite, ERIC, Dissertation Abstracts online, OSU periodicals online, and Central Washington University CATTRAX online database. Keywords used for searches included, but were not limited to: crew resource management, cockpit resource management, task management, cockpit task management, concurrent task management, multitasking, resource management, workload management, attention management, task prioritization, task switching, task shedding, aeronautical decision making, aviation Human Factors, human error, cognitive memory, learning theory, cognitive learning theory, social cognitive theory, mastery learning, overlearning, and constructivism, line oriented flight training, and the names of various authors.

Primary and secondary references were also obtained by hand searches of library collections and periodicals at Oregon State University Valley Library, Portland State University Library, the University of Central Florida Library, and the Central Washington University Library. Searches included references in the fields of psychology, aviation, and education.

2.1 Human Factors in Aviation

The concepts of Human Factors, ergonomics, and human error are intertwined in the literature such that it is difficult to find a concise definition of each. Glendon and McKenna (1995) noted that many writers consider the terms ergonomics and Human Factors to be synonymous and use them interchangeably to describe any subject that addresses the human-system interface. Hawkins (1993) stated:

Human Factors is about people. It is about people in their working and living environments. It is about their relationship with machines and equipment, with procedures and with the environment about them. And it is also about their relationship with other people. While all definitions are man-made and rarely carry the force of law, they are useful in guiding enlightened discussion and crystallising professional activity. They should not be too restrictive, however, and should allow
for development and new knowledge. The most appropriate definition of the applied technology of Human Factors is that it is concerned to optimise the relationship between people and their activities by the systematic application of the human sciences, integrated within the framework of systems engineering. Its twin objectives can be seen as effectiveness of the system, which includes safety and efficiency, and well-being of the individual. (p. 20)

Discussion and study of aviation Human Factors includes physical systems, physiological criteria, and psychological issues with the ultimate goal being to manage error and maximize safety. Some elements include flight deck and cabin layout and design, systems and equipment, computer software, charts, checklists, and documents. Flight training manuals typically group Human Factors into elements such as pilot fatigue, flight planning, pilot in command responsibility, stress, fitness, workload management, communications, situational awareness, cockpit design, and medical issues as factors that directly affect the pilot or flight crew during flight. Additionally, other factors that originate from human states or interactions outside the cockpit, such as professional and organizational culture; life stress; personality and behavioral attributes; and cultural, political, and social environments play an important role in pilot performance and are covered extensively in the literature. In the broadest sense, aviation Human Factors refers to the concept of anything that relates to human behavior in the flight environment. ADM, and more specifically minimizing and mitigating task prioritization error in the cockpit, comprises the focus of this study and as such drove the literature review.

The following section gives a brief overview of the historical development of Human Factors and human error as they relate to aviation so that a constructive focus is maintained with regards to the scope of this study.
2.1.1 Historical Development

The study of Human Factors in science and technology is relatively young; its birth was a little more than a century ago (Hawkins, 1993; Reason, 1990). Originally centered on the physical relationship between humans and machines, Human Factors moved into the realm of psychology when studies done at the Hawthorne Works of Western Electric from 1924 to 1930 identified the influence of psychological factors not directly related to the work itself, such as motivation (Hawkins, 1993). Hawkins described the evolution of aviation Human Factors research and application:

Since the pioneering days of flying, optimising the role of man and integrating him in this complex working environment has come to involve more than simply physiology. In particular, it has come to be concerned with human behavior and performance; with decision-making and cognitive processes; with effective use of the equipment in all operating environments; with the design of controls and displays, and with flight deck and cabin layout; with communication and with the software aspects of computers, maps, charts, and documentation. It is also concerned with the refinement of staff selection, and training and checking, all of which require skilled Human Factors input...A shift of emphasis has clearly taken place from physiology toward psychology. (p. 21)

The dawn of commercial aviation was in the 1920’s; a time when machines were highly unreliable and accident rates enormous. A government study cited in Kern (2001) cited an accident rate of 286 per 100,000 flight hours, mostly due to mechanical or engine failure. That time represented huge losses compared to current commercial accident rates of less than 0.3 per 100,000 hours (NTSB, 2005). As aircraft became more reliable fewer accidents were caused by mechanical failure. By 1940, about 70% of aircraft accidents were attributed to human performance errors (Muller, 1940, quoted in Hawkins, 1993).

During the Second World War, it became apparent that technology was outstripping human ability to operate complex machines and systems with maximum effectiveness, and the study of Human Factors was accelerated in military applications. In the 1940s, Cambridge University in the United Kingdom constructed
a cockpit research simulator to address ergonomics issues in aircraft design, and that group worked closely with the Applied Psychology Unit in an attempt to analyze effects of human systems related to human error in flight (Hawkins, 1993).

Additionally, aviation psychology centers were initiated at Ohio State University and the University of Illinois, Urbana. However, the science of aviation Human Factors was still in its infancy and the safety record did not seem to improve much for the world's military. In 1951, a U.S. Air Force Inspector General report found that "poor organization, personnel errors, and poor teamwork" resulted in the majority of aircraft accidents, and that "the human element...is essential to reducing the accident rate." (USIG, Poor Teamwork as a Case of Aircraft Accidents, quoted in Kern, 2001, p. 6).

Additionally, during the 1960's the U.S. Navy contracted with Boeing Aircraft to study Human Factors related to pilot performance and apply it to cockpit ergonomics and pilot training (J. Premesalar, personal communication September 17, 2003).

Civil aviation was not faring any better than the military. By the 1970's the safety trend had not improved. Three out of four accidents were attributed to some kind of human failure (Helmreich, 1993; Kern, 2001; Reason, 1990). It was apparent that basic education in Human Factors was vital to aviation safety (Hawkins, 1993; Helmreich & Wilmelm, 1991; Kern, 2001; Reason, 1990). As a response to this critical situation, the International Air Transport Association (IATA) organized a Human Factors committee and in 1975 held a conference in Istanbul devoted to Human Factors. According to Hawkins (1993), the Istanbul conference acknowledged that something was amiss in aviation human performance and that a basic Human Factors educational gap existed in air transport. The conference participants called for urgent action from the industry (Hawkins, 1993), but tragically it was not until two Boeing 747's collided on the runway in a dense fog on the island of Tenerife a little over a year later, killing 583 people, that any action was taken. As a direct result of the findings of that accident, KLM Dutch Airlines initiated the first "Human Factors Awareness Course" for indoctrination of flight crew and staff (Hawkins, 1993; Helmreich & Fousee, 1993; Kern, 1998).
As if to reiterate the need for human tragedy to precipitate action, airlines in the United States did not follow suit until after the aforementioned 1978 United Airlines crash in Portland, Oregon. The accident report identified several Human Factors as the cause and advised "all air carriers to indoctrinate their crewmembers on the principles of flight deck resource management" (NTSB DCA79AA005, 1979, p. 30). Based on the findings of that report, United Airlines developed and implemented training in Human Factors under the name of Crew Leadership and Resource (CLR) Training.

The year 1976 marked another important milestone in aviation Human Factors recognition, research, and application. The Aviation Safety Reporting System (ASRS) was established jointly by the FAA and NASA. Pilots, flight crew, controllers, dispatchers, administrators, and others can make confidential reports of incidents, behaviors, mishaps, errors made, and errors avoided. Currently over 300,000 reports exist in the database (Aviation Safety Reporting System, 2004) and are available for analysis. In 1982, the UK established the similar Confidential Human Factors Incident Reporting Programme (CHIRP), and in 1986 Canada formed the Confidential Aviation Safety Reporting Programme (CASRP). Australia followed suit with the Confidential Aviation Incident Reporting (CAIR) in 1988, and Germany started the European Directive for a similar program in 1992. Together, study and analysis of reports from these systems provides a wealth of statistics to study Human Factors in aviation, including situations that resulted in effective management of Human Factors and human errors toward successful outcomes.

When designing a training program to reduce human error, which was the focus of this study, it is important to gain a better understanding of the factors related to human error in flight operations. The following section is a brief overview of human error as it relates to aviation mishaps. This will set the stage for further discussion of the more specific types of error related to task prioritization as discussed in section 2.4.
2.1.2 Human Error

The literature suggests errors are natural human phenomena and are an important evolutionary learning tool (Glendon & McKenna, 1995; Hawkins, 1993; Reason, 1990). Glendon and McKenna pointed out that errors are essential to human growth and survival, and that all humans commit errors, but in today's world there are situations that are critically intolerant of errors. Reason (1990) equated aviation with other industries such as medicine and nuclear power and deemed them unforgiving of errors: "Whereas in the more forgiving circumstances of everyday life, learning from one's mistakes is usually a beneficial process, in the control room of chemical or nuclear power plants, such educative experiences can have unacceptable consequences." (p. 183)

Often, after some aviation disaster or aircraft accident, the media will use the phrase "pilot error" as if to place the blame of the entire incident on one individual or one flight crew. This terminology was used for decades in aviation accident and incident reporting. However, the pilot or flight crew should not take all the blame. Research shows that human error in the cockpit is typically the culmination of a long chain of circumstances and events which include not only the active failure (immediate action or lack of action) of the flight crew but also "latent failures" within a system, an organization, or even a culture that affect the tragic sequence of events leading up to the accident (Glendon & McKenna, 1995; Helmreich, Wilhelm, Klinect, & Merritt 2001; Kern, 1998; Reason, 1990; Shappell & Weigmann, 2001; Wiegmann & Shappell, 2001).

Simply writing off aviation mishaps to "aircrew error" is a simplistic, if not naïve, approach to mishap causation. After all, it is well established that mishaps cannot be attributed to a single cause, or in most instances, even a single individual. Rather, accidents are the end result of a myriad of latent and active failures, only the last of which are the unsafe acts of the aircrew. (Shappell & Weigmann, 2001, p. 23)
The FAA uses the term "poor judgment chain" to identify this series of mistakes that may lead to an accident or incident:

Two basic principles generally associated with the creation of a poor judgment chain are: (1) one bad decision often leads to another; and (2) as a string of bad decisions grows, it reduces the number of subsequent alternatives for continued safe flight. (FAA, 1991, p. 7)

The error chain or latent errors can reflect factors such as illness, fatigue, and complacency as well as improper procedures or poor training. However, the error chain can be broken in some instances through education and training of flight crews to detect, recognize, and manage errors more effectively (FAA, 1991; Kern, 1998; Reason, 1990; Shappell & Weigmann, 2001; Wiegmann & Shappell, 2001).

2.1.3 Summary

Human error and training in Human Factors both play a critical role in aviation safety. Modern training programs in Human Factors and aeronautical decision making (ADM) exist in airline and military training and have only become a part of general aviation training in the past decade. Significant improvements have been made in the military and airline industries in the past 20 years through Human Factors training. This type of training, with emphasis on mitigating task prioritization errors, was the crux of the CTM short course conducted in this study.

2.2 Cognitive Theories of Multitasking Behavior

Any time humans must make decisions there is opportunity for error. Pilot multitasking is no exception. Several theories of multitasking behavior are posited in the literature. Wickens (1992, 2002) defined multitask performance as the allocation of differentiated cognitive resources among competing tasks. Raby and Wickens
(1994) explained it as an attempt to manage workload and to balance acceptable levels of performance with acceptable levels of cognitive stress. A similar presentation of multitask performance as a function of workload was given by O'Hare and Roscoe (1990), but they related performance to the ability of the pilot to time share between concurrent tasks. Kern (1998) discussed effective execution of cockpit tasks as a function of proper procedural discipline in prioritizing both inputs and tasks.

Multitasking has been recognized as a key element to successful performance in complex systems (O'Hare & Roscoe, 1990; Raby & Wickens, 1994; Wickens 1990, 1992, 2002) and studies of processes people use to attend to multiple concurrent tasks are prolific in the literature. In the mid 20th century, cognitive psychologists used computer metaphors to describe the brain and cognitive processes related to performing multiple tasks. Three basic theories of multitasking and task management that have evolved from research in cognitive psychology are 1) single channel theory, 2) single resource theory, and 3) multiple resource theory. Each of these is discussed below in the context of processes involved and the limitations they present to effective multitasking abilities.

2.2.1 Single Channel Theory

Early researchers concluded that information must be processed sequentially based on the time available to perform tasks and that there is an overall limit on human ability to handle information and perform associated tasks (Broadbent, 1958; Lindsay & Norman, 1972; Welford, 1952, 1967). This type of mental processing is called the "single channel bottleneck" or single channel theory (SCT), and it assumes that no parallel processing or timesharing can take place: two tasks cannot be performed concurrently, and one will be dropped until the other is completed (Moray & Rotenberg, 1986; Wickens, Dixon, & Chang, 2003). Since SCT predicts that tasks must be performed sequentially, the following summarizes the relationship of concurrent tasks to time available:
SCT has different manifestations. All versions of strict SCT predict that progress on information processing can only take place on one task at a time, and therefore the completion time for two tasks imposed concurrently will equal the sum of the completion times for each done alone. This concurrent completion time will increase to the extent that information for a second arriving task is closer in time to the initiation of the first arriving task. (Wickens, et al., 2003, p. 12)

Based on those time-limited models of mental capacity, studies done for the U.S. Navy in the 1960's focused on overcoming limitations by developing more efficient ways of attending to sequential tasks. Researchers first focused on the amount of time it took for a pilot to process a task. For example, studies done by the Boeing Company determined that it took an average of 3.9 seconds for the pilot to acknowledge course data, 1.8 seconds to check attitude and heading, 3.8 seconds to change course to the new heading, and 5.0 seconds to monitor systems status. Next, procedures were developed to maximize the relationship between information processing and task performance times so pilots performed tasks in a specific sequence based on priority, which optimized task performance (Boeing Company unclassified study; personal communication J. Premesalar, September 17, 2003). In a similar study conducted with U.S. Marine aviators, pilots were trained to acquire all necessary information from cockpit instruments before refocusing their attention outside. Pilots who had this training were much better at detecting external targets than those who were not trained, even several months after the training had occurred (Gabriel & Burrows, 1968).

2.2.2 Single Resource Theory

Other research revealed that experimental subjects who were asked to process simultaneous messages could recall some characteristics of the second message, such as whether a speaker was male or female (Lindsay & Norman, 1972). Lindsay and Norman postulated that some kind of filtering mechanism limited the overall capacity
to transfer incoming sensory information into working memory. This signified an important component to the evolution of cognitive theory called Single Resource Theory (SRT), which differs from SCT in that cognitive resources, rather than amount of time available, predict task interference and performance (Wickens et al., 2003). This theory posited that there is a single pool of cognitive resources available, but those resources are undifferentiated with regards to attention (Kahneman, 1973). According to SRT, when more than one task is performed, or when tasks become more difficult, this pool of resources become limited. However, motivation and subsequent mobilization of increased effort could overcome the penalties of increased task difficulty so that two tasks could be performed simultaneously, although task performance might be degraded in one or both tasks (O’Hare & Roscoe, 1990; Wickens et al., 2003). Moray (1967) determined that timesharing could occur in certain circumstances where humans had the ability to share cognitive resources between tasks, and one model enabled concurrent processing of tasks based on "mental effort" and attentional resources (Rolfe, 1973).

### 2.2.3 Multiple Resource Theory

An important concept that emerged from several studies was the notion of time sharing, or the ability to alternate between different sources of information. During initial training a pilot will spend the majority of time focusing on the primary task of learning to manipulate the controls to fly the aircraft. As a pilot gains skill and confidence, she will have more time to divide her attention to other tasks such as scanning for traffic, monitoring instruments, and assessing the status of current and future situations. However, further studies showed that ability to perform tasks concurrently and efficiently depended not only on time sharing ability but on the cognitive resources or processing demands imposed by the individual tasks (North, 1977; Wickens, 1980; Wickens, Vidulich, & Sandry-Garza, 1984). For example, monitoring the Electronic Flight Indicator System (EFIS) is a visual and spatial task,
whereas listening and responding to an air traffic control (ATC) clearance is an aural and verbal task. Because spatial and verbal tasks operate in distinctly different ways and take place in separate parts of the brain, there will be less conflict between those types of tasks, because they are not competing for the same mental resources. Wickens (1980) determined this to be an example of what he called multiple resource theory (MRT). According to MRT, human attention resources can be switched from one area to another, and some tasks can be carried out simultaneously more easily than others (i.e., tasks that do not compete for the same resources). However, if tasks are competing for the same type of resources, then task performance deteriorates as resources are reduced. If a pilot is trying to perform two visual tasks or two auditory tasks simultaneously, performance is more likely to deteriorate than if one of each type of task is being performed. A practical application of MRT lies in the design of cockpit systems such as voice activated control systems and auditory displays, which are less likely to interfere with the primarily visual spatial task of flying (Liu & Wickens, 1992; O'Hare & Roscoe, 1990; Wickens et al., 2003).

2.2.4 Summary

The way a pilot allocates cognitive resources to perform multiple concurrent tasks is an important aspect of multitasking theory. It is not within the scope of this study to delve more deeply into cognitive processes that might influence pilot behavior under all possible circumstances. However, participants in the study did make prioritization errors that could be described by each of the main theories (see section 5.2). Regardless of how a pilot processes information, there is still a great challenge presented by the need to properly prioritize tasks to ensure maximum safety and efficiency of the flight. The following section takes a closer look at some of the theory that focuses on ways in which a pilot can deal with limitations related to cognitive processing and use cognitive resources to manage attention, workload, and flight tasks.
2.3 Methods of Dealing with Cognitive Limitations

It is possible that pilot abilities to effectively process inputs, prioritize, and execute tasks includes individualized skills that vary greatly from one person to another. However, a certain level of multitasking ability is required to perform even at a basic level adequate to achieve pilot certification. That means pilots in general comprise a group that has been pre-selected to some extent based on their demonstrated ability to properly prioritize and execute tasks in the context of flight operations. The previous section discussed theories related to how pilots process inputs and prioritize them to help perform and prioritize tasks as well as some limitations presented by those cognitive processes. This section includes a closer look at concepts related to attention management, workload management, and task management as ways of dealing with those limitations. Also, the theory of Concurrent Task Management (CTM) is discussed in further detail, and questions are posed that warrant further investigation.

2.3.1 Attention Management

One approach to multitasking in the cockpit focuses on managing pilot attention with respect to inputs and to prioritization and execution of tasks. Kern (1998) put it this way:

Attention management is a very complex phenomenon involving both the conscious and subconscious. It keys off of pattern recognition, or the ability of the brain to make sense out of multiple inputs by arranging them to fit patterns it has seen before. Often in aviation, there is no pattern established in your memory banks for a new situation, and this can lead to severe task saturation and channelized attention, two of the grim reaper's favorite tools for use on aviators. In order to make sure that we have the necessary attention available to complete mandatory procedures, we must learn to manage our attention. (p. 90)
Kern (1998) pointed out that task saturation can occur as a result of two different situations. The first is information overload, where the brain's ability to comprehend is simply overwhelmed by the mass sensory input. The second occurs when a pilot fails to adequately prioritize inputs so that unwise time-sharing between important and unimportant tasks occurs. Shappell and Wiegmann (2001) noted that attention failures are linked to errors such as breakdown in visual scan patterns, task fixation, and even inadvertent activation of controls, such as that which lead to the tragic crash of Eastern Airlines Flight 401 into the Florida Everglades in December, 1972. The crew became fixated on a landing gear indicator light, and one of them bumped the control yoke causing the autopilot to initiate a descent. The crew failed to notice the airplane was in a descent and failed to investigate even when ATC queried them regarding their altitude. The flight crashed into the ground killing 94 passengers and 5 crewmembers (NTSB DCA73AZ005, 1972). Shappell and Wiegmann (2001) compared that accident to a driver who is in hurry, or daydreaming, and misses their exit. They added, "These are both examples of attention failures that are commonly occurring highly automated behavior. While at home or driving around town, these attention failures may merely be frustrating. However, in the air they can become catastrophic" (Shappell & Wiegmann, 2001. p. 63). Funk et al. (2003) defined this type of error, where the pilots or crew focused their attention on a less important task and omitted a more important one, as a CTM error.

Kern (1998) noted that to effectively prioritize inputs and actions "one key is to stay ahead of the aircraft and to use times of relatively low workload to accomplish future tasks" (p. 90). He described a second "indispensable survival tool for pilots when dealing with task saturation is a system for prioritization when the stuff hits the fan" (p. 91). Kern (1998) pointed out that pilots must not only be able to prioritize tasks, but to prioritize information and input to avoid time-sharing between important and unimportant tasks; failure to do so can result in channelized attention or task overload, which is a major cause of breakdown in procedural discipline. It follows that in order to effectively execute tasks with proper priority and avoid task saturation pilots must learn to manage their attention. Kern (1998) emphasized that procedural
discipline is the best solution for prioritization during busy times and that pilots should use the adage familiar to all flight students - "aviate, navigate, communicate" - as a hierarchy to assist with prioritization. Chappell (1998) amplified this concept with the following words:

From the very first flight lesson, we were taught to "aviate, navigate, communicate," in that order. To aviate, navigate, and communicate, you must be aware of the plane, the path, and the people (crew, passengers, dispatchers, and air traffic controllers). Not only do you need to monitor and evaluate these three things now, but also you need to anticipate what's going to happen in the future and consider contingencies. The current and future state of the plane, the path, and the people are the components of the plan. (pp. 249-250)

The aviate, navigate, communicate (ANC) prioritization scheme described by Kern (1998) and Chappell (1998) for assessing inputs and accomplishing tasks is universally taught in primary and advanced flight training (Jeppesen, 2003; Kershner, 1998; Machado, 2001, 2003; Thom, 1991) and will be the scheme used to define CTM errors and for training in this study.

As discussed by Kern (1998) flight training places strong emphasis on procedural discipline as paramount to workload and attention management; when sensory overload, interruptions, and distractions threaten flight safety, procedures may be all the pilot has to fall back on to prioritize their inputs, tasks, and actions.

2.3.2 Workload Management

The ability to prioritize tasks is closely related both to a pilot's ability to focus attention and their ability to manage workload (Kern, 1998; Wickens, 2002, Wickens et al., 2003). From their very first training, pilots are introduced to the concept of workload: "Effective workload management ensures that essential operations are accomplished by planning, prioritizing, and sequencing tasks to avoid work overload" (Jeppesen, 2003, p. 3-34). One way to approach the concept of workload is to help the student understand that the majority of accidents occur during periods of high
workload, which include takeoff and landing, and that sticking to the ANC task prioritization scheme is paramount to avoiding distractions during critical flight times (Chappell, 2001; FAA, 1999; Jeppesen, 2003; Kern, 2001). Additionally, students must gain insight into the role of distractions and how to deal with them as a component of managing their workload. Jeppesen (2002) recommends: "If students must perform multiple tasks at the same time, ensure their attention is not focused on one item too long to the exclusion of others" (p. 3-35).

From an aviation psychology perspective, workload is probably a variable concept depending on the ability of the pilot or crew and on their preparation and planning strategies and practices. With certain combinations of tasks, individuals differ in their ability to process simultaneous inputs (Braune & Wickens, 1986). Wickens (1992) determined that individuals have an optimal level of workload and that above or below that level both individual and composite task performance is diminished:

Mental workload can be described as the relationship between resource supply and task demand. If supply exceeds demand, then performance is constant. But if demand exceeds supply, then performance will decrease as the resource demand (workload) further increases. Each of the pilot's responsibilities impose a certain amount of demand. The question is how much supply the pilot has available to cope with that demand, and when the demand reaches a point where performance drops due to a lack of resources. (Wickens et al., 2003, p. 3)

While the ability to manage workload may be highly individual, but both practice and adherence to procedures can contribute to increased ability to manage workload effectively in the cockpit (Chappell, 2001; Kern, 1998).

2.3.3 Task Management

Rather than focusing on workload management, some studies have approached multitasking from the concept of task management, which is distinct from workload
management in that it entails managing discrete tasks rather than total workload (Raby & Wickens, 1994; Rogers, 1996; Schutte & Trujillo, 1996). Regardless of how workload is approached, the prioritization of tasks as an integral part of managing workload is widely recognized in the flight training industry. One of the most widely used pilot training manuals on the market (Jeppesen, 2003) lists task prioritization as an integral component of workload management in the very first chapter: The manual instructs student pilots that "As pilot in command, you are responsible for determining the order in which tasks should be accomplished so that items which are essential for safe operation of the aircraft are performed first" (p.1-59).

Proponents of task management and CTM theory posit that humans controlling complex systems such as aircraft engage in a high-level mental process that continuously prioritizes concurrent tasks and allocates resources to them (Funk et al., 2003). Raby and Wickens (1994) conducted experimental studies of student pilots flying simulators in various flight situations. In their study, 30 student pilots flew three simulated landing approaches under low, medium, and high workload scenarios. They investigated how the pilots decided to prioritize tasks and shed tasks once they were completed. Raby and Wickens (1994) determined that people adapt to high workload periods by prioritizing tasks; the higher the priority, the closer the task was performed at the optimal time. Also, as workload increased, some pilots' performance on primary tasks (flying the airplane) diminished to the point of creating dangerous situations. A significant outcome of Raby and Wickens (1994) study was that individuals assume or shed tasks in order to maintain workload at a relatively constant level which varies with the individual. Pilots who were most successful were those who scheduled discrete tasks during periods of low workload (Raby & Wickens, 1994).

Raby and Wickens (1994) empirical results substantiate the positive results of proper flight training, because what they observed corresponded with what pilots are taught to do as a component of flight training. During FAA flight instructor training, strong emphasis is placed on task prioritization (FAA, 1999; Jeppesen, 2002). Flight instructors are reminded that “To manage workload, students must be taught how to
prioritize items" (Jeppesen, 2002, p. 3-34). However, that text notes that "Students must understand that priorities change as the situation changes" (Jeppesen, 2002, p. 3-34) and that one way to assure a student’s ability to apply that understanding is to help them focus on proper procedure, including scheduling of non-essential tasks during periods of low workload. Jeppesen (2002) noted that "To help students manage workload and avoid distractions, teach them to perform the majority of head-down tasks, such as reviewing charts or navigation logs during low workload periods" (p. 3-35). Pilots call this "staying ahead of the airplane."

Empirical results cited by Liu and Wickens (1992) and Wickens et al. (2003) showed that the type of task being performed affects task performance. According to those studies, task performance was degraded most when tasks being performed required similar types of mental processing (e.g., two visual tasks) and pilots were better at performing concurrent tasks that used different cognitive resources such as a visual and an auditory task. Liu and Wickens (1992) conducted experimental studies in which pilots were assigned a primarily visual task of tracking a course and then asked to perform either a spatial decision task (e.g., predicting the future position of a vector) or a verbal task, such as mental arithmetic. They determined that an inherently spatial visual scanning task produced more interference with a concurrent spatial task than with a concurrent verbal task: "tracking error, decision accuracy, and workload all suffered more when both tasks involved spatial activities" (p. 141). Wickens et al. (2003) stated that increased perceptual competition disrupts a cognitive task more than a motor task:

Primary task performance can suffer immensely while a pilot focuses most, or all, of her attention on dealing with the secondary task. When designing a system that requires a cognitively challenging secondary task, it is important to determine exactly how that secondary task will affect performance in other concurrent tasks. (p. 8)

In order to improve task prioritization Wickens (2002) suggested using the aviate, navigate, communicate hierarchy and added operation of systems as an additional task placing it last in the hierarchy (defined as ANCS). Wickens (2002)
noted that the extent to which the hierarchy is maintained when an ongoing task is interrupted by an incoming task depends on the type of interrupting task:

Some evidence suggests that auditory tasks low on the ANCS hierarchy, and particularly auditory communication tasks, tend to be both more interrupting and less interruptible than tasks with a higher priority (e.g. navigation). Studies comparing better and more poorly performing pilots have indicated that better multitask performance results from rapid switching between tasks (Wickens, 1999).
(Wickens, 2002, p. 132)

This empirical observation corroborates what flight instructors have known for decades: pilots will typically place communications first on the list of tasks even when they know it should be the lowest priority. This investigator has observed the phenomenon hundreds of times while giving flight instruction. Although based on anecdotal knowledge rather than scientific research, flight instructors agree it is one of the most common and most potentially dangerous task prioritization errors. Popular aviation training magazines, such as Flight Training, which targets student pilots and flight instructors, repeatedly publish articles addressing this issue of task misprioritization. For example, Miller (2003) wrote: "You may not have heard of Marconi's law. Named somewhat facetiously for Guglielmo Marconi, who transmitted the first wireless message in 1895, it says, 'fly the airplane, not the radio!'" (p. 38).

Pilots should focus on flying the aircraft (which in most cases includes operating the systems) as top priority, followed by navigating, dealing with emergencies, and finally letting someone know about it if they have time. To quote a popular aviation anecdote: "Flying the airplane is more important than radioing your plight to a person on the ground incapable of understanding or doing anything about it" (anonymous).

Wickens (2002) described the ability of a pilot to perform the complex tasks related to flying an airplane as "an interlinking set of cognitive phenomena relating to awareness, aircraft control, attention, mental resources, and strategic task management" (p. 132). The challenge of training pilots to better prioritize tasks should take on a holistic approach, and any CTM training criteria should be well founded in current aviation flight training practices and procedures.
2.3.4 Training in Attention, Workload, and Task Management

It is important to note that some of the studies cited show evidence that training can improve task performance (Gabriel & Burrows, 1968; Premesalar, 1969). O'Hare and Roscoe (1990) stated that "It is possible to perform certain nonconflicting tasks concurrently without decrement to either, and workload studies have shown that this can indeed be the case" (p. 193). However, they also pointed out that experts' and novices' performance varies significantly and cited experimental results that indicated extensive practice is essential to improving the ability to time share between tasks and perform multiple tasks concurrently (p. 197). O'Hare and Roscoe wrote:

Expert performance of tasks carried out concurrently appears to become both effortless and accurate compared to novice performance...At the outset, processing is slow, effortful, and largely sequential. With practice, performance becomes faster with elements processed in parallel. These two styles of processing are referred to as controlled and automatic, respectively. Provided there are consistent elements in the task, such as following exactly the same sequence of actions in response to an engine fire warning, extensive practice will lead to more automatic processing. (p. 196)

This concept of practice, along with procedures training, was an important aspect of the CTM course analyzed in this study.

Although general aviation flight training uses the aviate, navigate, communicate (ANC) prioritization hierarchy (Chappell, 1998; Jeppesen, 2001, 2003a, 2003b; Kern, 1998; Kershener, 1998; Machado, 2001, 2003; Thom, 1991), Funk et al., (2003) chose to represent the task prioritization hierarchy as "aviate, navigate, communicate, operate systems (ANCS)." Wickens (2002) also used a similar prioritization scheme in which operation of systems was placed last in the hierarchy but did not specify where that scheme originated. The addition of systems as a separate and final task is not consistent with prioritization schemes taught in general aviation. Most authors and flight training syllabi treat operation of systems as a component of the aviate task. For example, Thom (1991) defined the aviate task as
“flying the airplane (HDG, altitude, airspeed, engine, systems, checking HI against compass)” (p. 18-9). Jeppesen (2003) stated:

As you manage workload, you must prioritize items. For example, if you need to perform a go-around, adding power, gaining airspeed, and properly configuring the airplane are priorities. Informing the tower of your balked landing should be accomplished only after these tasks are completed. (p. 10-35)

Configuring the airplane includes position of landing gear, flaps, and other lift or drag devices. In the case of a go-around, proper aircraft configuration includes retracting the landing gear and setting the flaps to takeoff position, which are both high priority and considered aviate tasks because they directly affect the aerodynamic efficiency of the aircraft in a critical situation close to the ground. To a pilot, putting operations of systems as a lowest priority, although it might make sense in theory, could seem at odds with what they have been taught. In most situations, particularly emergencies, operation of systems is regarded as part of the "aviate" task but could have a lower or higher priority than other aviate tasks, depending on the situation. Additionally, pilots are taught early on that priorities change as the situation changes (Jeppesen, 2003).

To determine which has higher priority, pilots usually refer to standard operating procedures (SOPs) checklists, cockpit flow checks, mnemonic memory aiding devices, and training, with the assumption that by following those procedures they are conducting tasks in the proper sequence. A flow check is a series of tasks that are performed in a certain sequence based on the layout of the cockpit. Engine and systems controls, fuel selectors, switches, and other important items are positioned so the pilot can perform tasks in a certain sequence (for example left to right or up to down). The design of modern cockpits also follows a standard placement of many of those controls so pilots can transition from aircraft to aircraft more easily. Most procedures and checklists are carefully constructed so that pilots perform tasks in an exact sequence, often in the order of importance or highest to lowest priority. Placing operation of systems as a last priority does not fit with most standard procedures, checklists, and flow checks found in the industry. The problems arise when pilots are
confronted with a rare or complicated situation and must prioritize tasks when there are no standardized procedures available. These situations are not common, but ultimately comprise the most hazardous because of their abnormal nature.

The discrepancy between the ANCS scheme described by Wickens (2002) and Funk et al. (2003) and the ANC hierarchy used in flight training represents a difference between theory developed by aviation psychologists and engineers versus standard operating practice and procedure utilized by the FAA and other professional flight training curricula. This gap between theory and practice might be a tough one to close since many flight instructors may not be familiar with models of cognitive processing or CTM theory. Additionally, aviation psychologists and engineers may not be entirely familiar with flight training industry and flight school practices and procedures. However, since the current study hoped to inform flight training practices it used the three-part ANC hierarchy currently in use by the aviation industry as stated in various flight training handbooks and manuals.

Pilot training, especially in a professional program such as the university course from which participants in this study were drawn, is highly procedural in nature and the pilot will already have some procedure to fall back on for most normal and non-normal or emergency situations. Typically the “aviate” task consists of a series of tasks that are part of a procedure, a checklist, or a flow check that flight students and pilots have already practiced extensively. For example, on the takeoff roll in a single engine airplane or simulator the pilot will advance power, steer the aircraft with her feet, and look immediately at the following gauges as part of flying the plane: she will look first at the manifold pressure, tachometer, fuel flow, and cylinder head temperature gauges before even glancing at the airspeed indicator. Once she has established that the engine and associated systems are “go” (or aborted the takeoff if they are not), the procedures is to call out “airspeed alive” followed by calling out the correct speed for rotation and calling “rotate”.

After rotation, the pilot will establish the proper pitch attitude for climb and call out “positive rate, gear up” when there is not enough remaining runway on which to land and a positive rate of climb is established. After that she will commence an
initial climb checklist which usually consists of retracting flaps if they were used for takeoff, turning off fuel pumps and checking fuel flow indications, setting the manifold pressure (using the throttle) and the tachometer (using the propeller control) for the proper climb configuration, and maintaining either a constant airspeed or rate of climb using pitch control. The actual act of flying the plane (aviate task) is a combination of many items, including the proper operation of systems in the proper sequence.

Retracting the landing gear is not a last priority and is considered an aviate task because the position of the landing gear, as well as the forces caused by the retraction or extension of the gear, affect the aircraft aerodynamics. Retraction of the landing gear affects the amount of drag on the airplane (and thus available thrust) which affects the aircraft’s climb capability. That situation could be potentially dangerous if an engine were to fail when the airplane is close to the ground with a high positive pitch attitude. If the pilot put navigation or talking ahead of retracting the gear and flaps she would be misprioritizing tasks at a critical time.

Ultimately, prioritization of tasks should be based on which tasks are most critical at the time and priorities will change constantly as they did for the United Airlines captain that December day in Portland in 1978. Essentially, most normal and emergency procedures, checklists, and flow checks are set up so that when the pilot executes them she is conducting tasks in the proper sequence from highest to lowest priority. However, when a task is non-normal in nature it is difficult for any hierarchical scheme to stand up completely under close scrutiny. Funk commented that the ANCS hierarchy did not fit all situations and was sometimes problematic and pilots often find that the ANC hierarchy does not fit all situations (K. Funk, personal communication, May 12, 2003). Therefore, strict definition of any prioritization scheme may pose problems in real flight time due to the dynamic nature of the status and urgency of tasks, as described by Funk et. al. (2003), and lower priority tasks may take higher priority over time and visa versa.
2.3.5 Summary

A complete discussion of methods to effectively deal with cognitive limitations related to multitasking and task management would fill volumes and is beyond the scope of this literature review. Generally methods center on management of some combination of pilot attention, workload, and allocation of tasks. During initial training pilots are introduced to the aviate, navigate, communicate hierarchy as well as trained in the use of checklists, cockpit flow checks, and standard operating procedures. Those techniques are designed to help the pilot manage attention and workload and properly manage tasks more effectively through standardization of procedures. The training course used in this study was based on typical concepts and practices used for ADM training and is described in more detail in Appendix B. Before proceeding to description of the study, a more in-depth discussion of historical and current concurrent task management (CTM) research, CTM theory, and previous CTM training is presented.

2.4 Concurrent Task Management

As defined by Funk (1991) and Funk et al. (2003), Concurrent Task Management (CTM) entails the initiation, monitoring, prioritization, execution, and termination of multiple, concurrent tasks by pilots or flight crews.

Concurrent Task Management is the process by which pilots selectively attend to tasks so as to safely and effectively complete the mission. CTM is critical to flight safety in that attention to a less important, less urgent, or already more satisfactory task can result in the dangerous interruption of a more import, more urgent, or less satisfactory task, and thereby compromise safety of flight. A significant number of aircraft incidents and accidents can be attributed to task management errors. Therefore, improving CTM performance should be a high priority to those concerned with improving the safety of the air transportation system. (Funk, et al., 2003, p. 58)
The preceding definition and insight into applicability of effective CTM has been the focus of an ongoing NASA funded research project at Oregon State University that has resulted in several doctoral and master's thesis and extensive research into multiple aspects of CTM (K. Funk, personal communication, 2003). Those studies will be discussed in various sections of this part of the literature review. First, factors that have been determined to affect CTM will be explored, followed by a brief review of technological aids to CTM performance, and finally a discussion of previous CTM training experiments. Funk, et al.'s (2003) theory will be put into perspective with regards to the proposed study, particularly the factors that affect CTM performance. The complete CTM theory developed by Funk, et al. (2003) is available on request. An important part of the discussion relates to previous training and experimental designs used by Funk, et al.'s group as discussed in the last part of this section.

2.4.1 Task Prioritization Factors

CTM is an ongoing process by which pilots initiate new tasks, monitor ongoing tasks, selectively prioritize tasks, and terminate, or shed tasks deemed less important or that have been completed (Funk, 1991).

CTM is not new; in fact, pilots have always done it. CTM is a cognitive function that is intuitively well understood by pilots and almost always performed satisfactorily. However, there are many documented instances in which tasks were not managed properly, resulting in an aircraft incident or accident (Chou et al, 1996). Often, during critical phases of flight, this form of human error results in minor regulations violations or unsafe conditions that are rectified before a more serious situation develops. However, the consequences of improper CTM can be a catastrophic event resulting in many fatalities and loss of the aircraft. (Funk et al., 2003, p. 9)

One question that arises is: what are the factors that affect how a pilot chooses to prioritize tasks, and how does he decide which tasks are more important? Adams,
Tenney, and Pew (1991) observed that a busy pilot must always be aware of the tasks that need to be performed and in what order they must be performed. Because the number of concurrent tasks is so great and because in many cases each task is critically important to flight safety, pilots are accustomed to relying on checklists, flow checks, standard operating procedures, and in many cases mnemonic memory aids to help sort out task priorities. Kern (1998) and Chappell (1998) both emphasized that a critical factor is for the pilot to stay ahead of the aircraft and use times of relatively low workload to accomplish future tasks, which requires a high level of discipline. Kern (1998) emphasized that pilots must also have a system for task prioritization, such as the aviate, navigate, communicate hierarchy, proper procedural discipline, or use of the numerous mnemonic devices used in the flying industry.

An example of a mnemonic memory device that is used to execute tasks in a specific order based on highest to lowest priority is the “six Ts” commonly used to execute tasks when commencing an instrument approach procedure (some pilots use as few as five or as many as 10 Ts depending on the aircraft and the situation). The six Ts, in order of priority, are “time, turn, tires, throttle, track, talk.” It is critical to perform the Ts in the right sequence, because they are set up to properly prioritize tasks. Time is the most important task and is done first, because the missed approach point is usually defined by time (even on a precision approach it is important to time the approach in case the vertical glideslope instrument or ground based transmitter fails during the approach so the pilot can still fly a timed localizer only approach). Turn means to turn the aircraft to the new heading, if required, and tires means to extend the landing gear, thus adding drag and changing the aerodynamic configuration of the aircraft. The throttle is reduced after tires down. If the pilot set the throttle before extending the gear, she would have to reset it after the drag changed with extension of the gear. Track means following the localizer/glideslope or other non-precision course, and talk is always last.

In addition to mnemonic memory aids, pilots also use cockpit flow checks that consist of specific tasks performed in a specific sequence that is laid out according to the physical configuration of the cockpit. Usually tasks are performed in a left to right
or up to down motion. Thus, memory aids, procedures, checklists, and flow checks are the most common and highly effective way for pilots to prioritize tasks in normal and non-normal situations. However, when a pilot must rapidly switch between tasks, or when unexpected events require actions that are not a part of standard or emergency checklists and procedures, some highly cognitive tasks, such as maintaining situational awareness, cannot be easily codified in checklists and procedures (Wickens, 2002).

Although flight training manuals and pilot training courses contain numerous examples of procedures and task prioritization devices, formal literature provides only a few tentative answers to the question of what task prioritization factors pilots use (Funk et al., 2003). Therefore, the Oregon State University group conducted two separate experimental studies to better identify what factors pilots used to prioritize tasks (Colvin, 2000). In the first study eight male airline pilots flew two different instrument approach procedures in a part-task computer-based simulator developed by NASA (it was not an actual flight simulator as defined by the FAA). The simulator represented a generic, twin-engine turbojet with an advanced autoflight system, Electronic Flight Information System (EFIS), Horizontal Situation Indicator (HSI), and Engine Indication and Crew Alerting System (EICAS).

The pilots were given challenges throughout the flights that forced them to have to select between multiple tasks at a given time. Two different methods were used to elicit pilot responses as to why they chose to give a task a higher priority. The first method was intrusive, in which the simulator was stopped, the pilot asked as to why he had just prioritized a task, and then the flight continued. The other method was a retrospective technique in which the pilot flew the entire scenario, then reviewed a video of the flight and commented on why he prioritized tasks. Data gathered in the cognitive interviews with the pilots was used to determine factors that affected pilot task prioritization in the pilot's own words. No statistically significant differences existed between results from the intrusive and retrospective techniques.

Findings from that study (Figure 2.1) indicate that task status (e.g., how well the task is being performed) and procedure (e.g. if it is part of standard checklists,
operating procedures, or regulations) were the greatest determining factors identified by the pilots as to why they gave a task high priority.

Figure 2.1. Graph of task prioritization factors reprinted with permission from Funk et al. (2003).

In the second experiment, the task prioritization factors were reduced to six, based on criteria from the first experiments. Those factors were:

- Consistency of the task with standard procedures
- Importance of the task with respect to the mission goal
- Salience of stimuli related to the task
- Status of the task
- Perceived time and/or effort required to perform the task
- Urgency of the task
In Funk et al.'s (2003) second experiment, task performance data were collected and analyzed with the objective of identifying a relationship between task performance and prioritization factors used by pilots. Data were collected from four simulator flights, all variations of the same arrival scenario. The differences between flights were in the placement of a Challenge Probe Point (CPP) within a particular scenario. A CPP was an operational situation where up to six tasks could become active at the same instant and the pilot had to prioritize and perform the tasks. Funk et al. (2003) used the retrospective interview technique since it was less intrusive and was determined statistically equivalent to the intrusive technique. While reviewing a video of each flight, the pilot answered questions and filled out a questionnaire regarding each of the six CPPs. Pilots rated all six tasks as to what factors (of the six) affected the order in which they performed each task. Regarding the data analysis and results, the experimenters stated:

Our knowledge elicitation techniques required the pilots to identify a very specific order in which they performed tasks. We acknowledge that expert pilots may have the ability to perform multiple, concurrent tasks and we do not attempt to refute that point in this work. However for these studies, we assume that there is a strict prioritization of tasks with a single task emerging as the most prominent task, occupying a significant amount of the pilot's attention. (Funk et al., 2003, p. 12)

That discussion is consistent with contemporary flight training theory and practice, since many procedures entail breaking down complex tasks into their simple components and practicing them individually before synthesizing them and performing them simultaneously. The second study conducted by Funk et al. (2003) corroborated procedure as being the main reason pilots choose to prioritize tasks. "Over the entire experiment, procedure emerged as the factor most agreed to by pilots for use in task prioritization, followed in descending order by salience, importance, time/effort, urgency and finally status" (Funk et al., 2003, p. 16).

In addition to the experiments to determine task factors just discussed, Funk et al. (2003) designed and tested a model of task performance that confirmed the roles of task importance, status, and urgency as prioritization factors. They also conducted
experiments to determine the relationship of CTM performance to other cognitive processes and concluded:

Our preliminary findings indicate that cognitive abilities, at least as measured by the tests we used, are not good predictors of concurrent task management performance. CTM is a complex process that cannot be explained in terms of working memory, simple decision-making, or verbal intelligence alone. These results are not consistent with our expectations. (Funk, et al., 2003, p. 24)

In summary, the overall results of Funk et al.'s experiments corroborate Kern's (1998) and Chappell's (1998) statements that procedural discipline is paramount to a pilot's ability to manage attention, deal with task saturation, maintain situational awareness, and prioritize tasks. Additionally, all flight operations, and especially flight training, are highly procedural in nature. Therefore, it seems likely that most pilots would prioritize tasks based on procedures. That was an important in the design of the training course used in this study, which emphasized that in order to improve task prioritization pilots must improve procedural discipline.

2.4.2 Technological Aiding of CTM in the Cockpit

The Cockpit Task Management System (CTMS), described by Funk and Kim (1995), consisted of software that computed and displayed a list of up-coming tasks and in-progress tasks and recommended which tasks to attend to first. Experiments conducted by Funk and Kim (1995) showed that the CTMS reduced misprioritizations by 41% and reduced the number of incomplete tasks (i.e., those interrupted but not resumed to completion) by 82%.

Another such training aid called the AgendaManager (AMgr) was designed and tested by Funk and Braune (1999). The AMgr used a speech recognition system to determine pilot goals, then software monitored the progress of each task and transferred the information to a display in the cockpit which showed higher priority tasks with greater prominence. Empirical data from Funk and Braune's study showed
that pilots' task prioritization improved by almost 30 percent using the AMgr over conventional displays, and their ability to keep all concurrent tasks at a satisfactory level improved from 52% to 65% of the total time. However, Funk et al. (2003) listed several challenges to developing and employing CTM aiding systems in the cockpit:

Both aiding systems, CTMS and AMgr, were effective in part-task simulator evaluations and therefore suggest that aiding may be an effective way of improving CTM performance in an operational environment. But several challenges require that other avenues be explored before any full-scale development of a CTM aiding system be undertaken. First, the speech recognition technology on which the AMgr depends is not yet reliable enough for use in the demanding cockpit acoustic environment. Second, the complexity of the software needed to implement a fully operational CTM aid would be at least as great as that of a modern flight management system, making, third, the certification of such a device very challenging. Fourth, systems such as the CTMS or AMgr require advanced avionics systems, yet many older generation aircraft, lacking such equipment, will remain in service for a very long time, perhaps decades. (Funk et al. 2003, p. 7)

Based on their findings, Funk's group determined that a more reasonable solution to improving CTM would be to develop training programs. A successful CTM training program would not entail development and testing of expensive new equipment and software, could be implemented immediately, and could be used in older aircraft. In an experimental study by Schutte and Trujillo (1996), pilots flew simulated flights in which aircraft systems failures dictated diversion and landing at an alternate airport. Schutte and Trujillo found that pilots who used effective CTM strategies outperformed those who did not and concluded that CTM training could improve pilot performance and should be incorporated into pilot training.
2.4.3 Previous CTM Training Studies

Two training studies were conducted for Funk et al.'s (2003) group, and experimental results from both studies were inconclusive with regards to the effectiveness of CTM training. This investigator visited the test site three times in 2002 and 2003 and had numerous conversations and written correspondences with Kenneth Funk and one of the other researchers (Nicolalde). Based on those observations and discussions, it was mutually agreed that the inconclusive results were probably due to several factors, including a small participant sample size, the low fidelity simulator platform used, and the training course itself. The final report of the training studies and recommendations for further research (Funk et al., 2003) reflects those observations (Funk, personal communication, May, 2003) and is outlined below.

In the first CTM training study (Bishara, 2002) twelve general aviation pilots of varying age and experience were randomly assigned to one of three groups, with four pilots in each group. The groups consisted of a control group and two experimental groups that received CTM training. Each of the three groups flew a pretest and posttest simulated flight on the FTD, and the training groups received training in the interim. The control group took a break between the flights. The age range of participants was from 20 to 72, and although Bishara stated the data were gathered, no data appeared regarding each pilot's flight ratings or experience. Thus, there is no way of knowing from that report how the pilots' previous experience influenced the experimental results. Funk et al. (2003) acknowledged this in their final analysis:

A particular source of unwanted variability might have been among the participants themselves. Although seemingly equivalent in qualifications, participants reflected a wide range of flying experience. For example, some of the participants were flight instructors and some were not, and we did not control for that attribute or use it in assigning participants to groups. (p. 57)

The participants were each introduced to the experimental platform, which consisted of a personal computer based flight simulator program with a flight yoke,
pedals, and a mouse that controlled systems virtually, and each flew a practice flight before the experimental flights. The program used was Microsoft (MS) Flight Simulator 2000, a popular computer game that is not FAA approved as a flight simulator PCTAD, or FTD, and the flight controls used were not approved for training. Furthermore, use of a mouse to virtually control the systems did not correlate to real-world aircraft controls. Many of the cues used to pilot an aircraft are kinesthetic, based on the "feel" of the flight controls, thus when a pilot cannot feel the controls, she has no way to gather feedback to help with assessing her performance.

It is possible that some of the task prioritization errors recorded in that experiment could have been errors due to poor aircraft control as a function of the pilots' unfamiliarity with the MS flight simulator game or with using the computer mouse to control systems. For example, in Bishara's (2002) study, if a pilot did not level off from a climb within 200 feet, it was considered a task prioritization error. However, it could have been due to the unrealistic feel of the controls of the computer and the pilot's lack of proper visual or kinesthetic feedback. This also means that the differences in performance between participants could have been attributed to differences in their ability to learn the MS Flight Simulator program rather than a difference in performance due to training. Ken Funk agreed to the importance of those issues (personal communication, April 18, 2003) and referred to the problematic nature of using that experimental platform in the final report of their study:

Besides the yoke, throttle, and pedals, which did not provide realistic force feedback, our simulator used virtual, mouse-operated controls on the instrument panel. There were no motion cues. Combined, these deficiencies in realism may have presented enough of a cognitive challenge to the participants that their responses were not representative of how they would have performed in a real airplane. (Funk et al., 2003, p. 57)

After the familiarization flight, and the first experimental flight, the two training groups each received six pages of notes covering various CTM topics, and "The experimenter read the lecture training material out loud while the participants followed along with their own copies" (Bishara, 2002, p. 64). One group was also
introduced to a mnemonic training aid, and then both training groups flew a 30-minute practice flight before flying the second experimental flight. All of the pre- and post-training flights, training, and practice took place during a single session. Based on the description, this investigator questions whether the pilots actually received any training at all. As the FAA pointed out:

The instructor cannot assume that students remember something just because they were in the classroom, shop, or airplane when the instructor presented the material. Neither can the instructor assume that the students can apply what they know because they can quote the correct answer verbatim. For students to learn, they need to react and respond, perhaps outwardly, perhaps only inwardly, emotionally, or intellectually. But if learning is a process of changing behavior, clearly that process must be an active one. (FAA, 1999; p. 1-3)

Additionally, since learning is defined as a change in behavior as a result of experience (FAA, 1999; Schunk, 2003), the student must have adequate time to process information and practice learned skills to achieve the highest levels of learning. Section 2.5 describes in more detail the FAA and industry approved practices as well as learning theories related to flight training. Any flight training must be thoroughly grounded in such theory and practice to meet FAA and industry standards. During discussions with Ken Funk, he acknowledged this as an important factor in producing the inconclusive results and added the following to the report:

Perhaps a more likely...possible explanation is that our training materials simply were not very good. We are human factors engineers and aviation psychologists – not training specialists – and the training materials were constructed from good information but perhaps we attempted to present them using less than satisfactory training techniques. Perhaps – as we believe – CTM is trainable, but the development of effective CTM training materials is non-trivial. (p. 57)

The second study conducted by Funk et al.'s (2003) group used the same experimental design and training. However, the sample group was larger (15 participants) and rather than being randomly assigned to groups, they were prescreened and assigned to groups based on their performance on verbal and spatial intelligence tests and a 15-20 minute flight in Microsoft Flight Simulator 2002.
Although the design was intended to control for differences in pilot performance or familiarity with the flight simulator program, the results were again inconclusive. Funk et al. (2003) concluded that CTM training should be viable, but that their inconclusive results were probably due to the criteria just discussed. Based on those criteria, Funk et al. suggested the following:

The general process we developed for conducting CTM training research (theory application, scenario development, use of the simulator) should be used to continue CTM training research. The process described, above, especially that of Experiment 2, described in Chapter 8, should be used, with the following changes.

1. CTM training materials should be developed from CTM theory, but with the assistance of a qualified aviation training specialist.
2. A more homogeneous pilot population should be used as participants (e.g., a group of student pilots at about the same stage in training or a group of military aviators with more similar qualifications).
3. A larger group of participants should be used.
4. A flight simulator with more fidelity should be used.
5. More time should be given for simulator familiarization.
6. More powerful experimental designs should be considered.
7. Greater care should be taken in training delivery and experimental procedure so as to reduce unwanted variability. (Funk, et al., 2003, p. 59)

2.4.4 Summary

Concurrent Task Management as defined by Funk, et al.'s (2003) theory contributes a significant element to the description of human error as it relates to ADM and to aircraft accidents and incidents. Based on Funk et al.'s (2003) theory, including factors that affect CTM, this study addressed the limitations of that study and introduced additional modifications and changes to the experimental design and the task prioritization process based on standardized flight training procedures and learning theories appropriate to flight training. Specifically, criteria mentioned by Funk et. al. (2003) were addressed as follows:
The CTM training materials and training course were developed, implemented, and conducted by an FAA certificated and highly experienced flight instructor.

Participants were selected from a relatively homogeneous pilot population consisting of flight students from a single university program undergoing standardized flight training. The pilots averaged 190 total flight hours, and 56 logged instrument hours.

An FAA approved FTD was used.

A larger group of participants was used (N = 27 for this study).

Participants were all familiar with the FTD used in the study and previous flight time on the FTD used for the study averaged 15 hours.

Additionally, pilots in this study had more time between the pretest and posttest flights so that the training course could address longer term learning effects.

This study hopes to contribute to knowledge surrounding the question as to whether CTM performance can be improved through training. A more detailed discussion of factors that affected experimental reliability and validity is presented in Chapter 3.

2.5 Learning Theories Relevant to Flight Instruction

The design and development of any training course to improve pilot CTM performance must be grounded in sound teaching and learning theories appropriate to the type of training being delivered. To qualify as FAA approved training, all flight and ground instruction, including simulator training, must be conducted by certified flight instructors (CFIs) certificated under 14 CFR Part 61 Subpart H or Part 141 Appendix F. The reason for this is that instructors must have the depth and breadth of knowledge at a minimum of commercial pilot level, as well as being trained and tested in teaching and learning theory, concepts, instructional design and delivery, and practices as determined appropriate by the FAA.
The FAA bases its flight instructor training on the concepts of social learning, reciprocal teaching, constructivism, mastery learning, and the integration of cognitive, affective, and psychomotor domains of learning (FAA, 1999; Jeppesen, 2002). Therefore, these learning theories and concepts are appropriate to the design and delivery of a CTM training course. Theories relative to major concepts outlined by the FAA are introduced, then a more comprehensive development of learning theory appropriate to the proposed training course is discussed in the following section.

Briefly, social learning theories share the concept that humans learn through interaction with others and with their environment (Bandura, 1977; 1986, 1993; Schunk, 2003; Vygotsky, 1978, 1979). Thus, social learning is affected by the culture in which the individual is enmeshed and cognitive development results from shared experiences and interactions with individuals or groups that include both instructors and more competent peers. The FAA advocates a learning environment in which the instructor builds a trusting relationship with the student and utilizes concepts of social learning to enhance the student's progress throughout training (FAA, 1999; Jeppesen, 2002). The strong emphasis placed by the FAA on building trust is critical, because the student understands they are literally placing their life in the hands of the instructor during flight training. If the student does not trust the instructor then learning can be compromised. When instructors work to build trust and guide students through their learning in a secure environment they are practicing the concepts of mentoring as described by Daloz (1986).

Flight training practices outlined by the FAA also emphasize reciprocal teaching (FAA, 1999; Jeppesen, 2002). Reciprocal teaching is an applied teaching method based on Vygotsky's (1978) theory of social cultural learning. It is an interactive process in which the teacher poses questions or models a skill and as the process continues the student takes turns being the teacher (Prickel, 2002). The FAA incorporates reciprocal teaching into all their training syllabi through the "instructor does, instructor tells," "instructor does, student tells," and "student does, student tells" sequence of training (FAA, 1999; Jeppesen, 2002).
The concept of constructivism posits that students construct their own knowledge through an active learning process based on existing beliefs and experiences (Schunk, 2003). Flight instructors are taught:

Constructivism is based upon the idea that learners construct knowledge through the process of discovery as they experience events and actively seek to understand their environment. To employ Constructivism, your role shifts from the transmitter of information to the creator of experiences. (Jeppesen, 2002, p. 1-11)

Mastery learning, described as a major component of social cognitive learning (Bandura, 1977; 1986) is a process that incorporates goal setting in small increments combined with continuous feedback, correction, and enrichment to assure that the learner masters concepts and skills before progressing to the next level (Bloom, 1974; Burton, Brown, & Fischer, 1984). Mastery learning is especially effective when learning extremely complex skills, as described by Burton et. al. (1985):

The student is exposed to a sequence of environments (microworlds) in which his tasks become increasingly complex. The purpose of an individual microworld is to provide the student with a task that he can perform successfully using a simplified version of the final skill that is that goal. This allows the student to focus on and master one aspect of the skill in a context that requires related subskills. As a result, the student learns when to use the skill as well as how to use it. Thus the purpose of the sequence is to evolve the simplified skills toward the goal skill. (p. 139)

Mastery learning is a primary technique recommended by the FAA and is the principle method of instruction incorporated into the building block method of integrated flight instruction training courses (FAA, 1999; Jeppesen, 2002). Instructional scaffolding (Bruning, Schraw, & Ronning, 1995) is a related technique and involves a process by which the teacher controls the number of tasks to be learned and, based on the student’s progress, introduces the next set of tasks or skills until the student is able to master those skills and move on to the next level.

Some flight instructors advocate continued repetitive practice of a concept, skill, or maneuver even after the student has mastered the ability to perform to completion standards. Continuing to rehearse information after the initial learning is
accomplished is called overlearning (Krueger, 1929). Overlearning is a technique that has been widely used, especially when learning highly intricate or technical skills such as playing a musical instrument or learning a language, and has been tested often in the literature (Driskell, Willis, & Cooper, 1992). In a meta-analysis of 15 experimental studies that investigated the effectiveness of overlearning Driskell et. al. (1992) found that overlearning greatly increased performance, but the increase usually disappeared in a short period of time, typically between one and three weeks.

Recent research questions the validity of overlearning as an effective strategy (Rohrer, Taylor, Pashler, Cepeda, & Wixted, in press). In two separate experimental studies they discovered that the boost in learning certain memory recall tasks mostly disappeared within nine weeks. Additionally, overlearning was highly inefficient; a quadrupling of study time produced far less than a doubling of recall rate when subjects were tested one week later (Rohrer et. al., in press). The training course used in this study did not incorporate a significant amount of repetitive practice but did include exercises incorporated into the training course in which participants analyzed their performance during their normal flight times. Because the training for this study was meant to target a pilot’s working memory, overlearning, although possibly valid for short term retention, was not emphasized. Additionally, overlearning strategies are typically used in flight training for rote memorization tasks and may be most effective at those lowest levels of learning, but not for mastery learning or scaffolded instruction necessary to teach complex skills such as those involved in learning to fly.

Flight training is a kinesthetic, visual, cognitive, and often emotional learning experience. Flight students and pilots bring expectations, doubts, and often fears with them into the cockpit. Flight training is also highly performance-based, and expected levels of performance are clearly delineated at the onset of training. The FAA Flight Instructor Handbook noted that it is imperative for flight instructors to design training curricula that clearly state the performance-based objectives being used and that students must be aware from the outset what those objectives are (FAA, 1999).

The FAA flight instructor literature stresses that all aviation-related training must go beyond rote levels of learning and that students must achieve higher levels of
experience, application, and insight in each of three learning domains – cognitive, affective, and psychomotor (FAA, 1999). The cognitive domain involves knowledge and thought processes, and highest levels include application, analysis, synthesis, and evaluation of those processes. Students will exhibit higher levels of learning in the psychomotor domain through positive response, adaptation, and forming new movement patterns. Interestingly, the FAA noted that "The affective domain may be the least understood, and in many ways, the most important of the learning domains" (FAA, 1999, p. 1-11). Thus, effective design and delivery of any aviation training course must pay attention to how students are involved emotionally in the learning process. Such a course must also integrate educational objectives that stimulate appropriate mental activity, emotional states of mind, and demonstration of skills at the highest levels possible in each of the three learning domains. That statement brings us back to the underlying themes of social cognitive learning theory as described in the following section.

2.5.1 Learning Theory Foundations for CTM Training Course

Based on more than seven thousand hours as a flight and classroom instructor, this investigator has found that facilitation of mastery learning is a highly effective way to ensure positive transfer of learning to pilots and flight students. The techniques outlined by the FAA of building on what the flight student already knows and continuously challenging them to master skills incrementally to achieve higher levels of skill has proven effective. Additionally, as the FAA emphasizes in all training curricula, students need time to practice maneuvers and skills and to form their own understanding of relationships between theory and application (FAA, 1999; Jeppesen, 2002). Focusing on the flight student's strengths to build their confidence and self-motivation toward mastering concepts and skills allows them to be able to make mistakes (as long as it is safe) to help them better analyze and synthesize their understanding of those mistakes and create their own solutions. This combination of
techniques is founded in constructivism, mastery learning, and social learning theory. The CTM training course used in this study was specifically grounded in concepts of social cognitive learning theory (Bandura, 1977; 1986, 1993) and social cultural theory (Vygotsky, 1978).

Social cognitive theory focuses on the relationship between learning and motivation of the individual within a social context. Bandura (1977; 1986; 1993) defined social cognitive learning as based on two principle concepts regarding the learner: 1) self-directed learning, which is the ability of a learner to set and achieve realistic and attainable goals for themselves; and 2) self-efficacy, or the learner's positive belief in their own ability to learn and master a concept or task and to achieve the goals they have set. Both concepts rely on a high degree of motivation, which Bandura describes with reference to two distinct domains: external motivations (e.g., rewards, praise, reinforcement, recognition) and internal motivations (e.g., past experiences, personality, desires, goals, curiosity, choices, and persistence). In their flight instructor training curriculum the FAA references all of these criteria as important to flight training and identifies them within the broader scope of social learning theory (FAA, 1999; Jeppesen, 2002). In social cognitive theory, the internal domain dominates the learner's motivation. The flight student's concept of success is based on her own sense of personal achievement, belief in the significance of her contributions, and belief in her ability to achieve her goals. Those beliefs comprise the student’s self-efficacy (Bandura, 1993). Figure 2.2 illustrates that relationship.
Based on social cognitive learning theory, any strategy that enhances the individual's self-efficacy serves to increase her success in achieving her goals and learning objectives by giving her control over her own learning. The most significant strategies are:

- **Mastery learning** (Bloom, 1974; Bandura, 1993) and **reciprocal teaching** (Vygotsky, 1978) as previously discussed.
- **Modeling**, in which the flight student observes and interprets a behavior, then adopts that behavior if it has functional value or results in outcomes they value (Bandura, 1977; 1993; Decker & Nathan, 1985).
- **Social persuasion**, in which the teacher or flight instructor serves as the model for mastery learning by facilitating self-efficacy and self-directed learning. The learner is more likely to adopt a modeled behavior if the mentor is similar to the
learner and has admired status (Bandura, 1993). The FAA stresses this in its approach to flight instruction through emphasis on developing a common core of experience between flight instructor and student. Instructor must place the student’s learning as the primary lesson objective and maintain a professional, accommodating style while respecting and accepting the student (FAA, 1999).

Decker and Nathan (1985) based their behavior modeling concepts on the theoretical work of Bandura (1977) and restated his theory:

In more informal terms, in order for people to learn from behavior modeling training, they must observe what the model is doing, remember what the model did, do what the model has done, and later when the appropriate time comes, want to use what they have learned. (p. 4)

Decker and Nathan (1985) incorporated those four concepts based on Bandura’s theory into five strategies for behavior modeling training, which include 1) Modeling, or the presentation or display of a behavior; 2) retention, which includes the learner’s mentally practicing the behavior or coding it by writing it down or verbally describing it; 3) rehearsal, which includes the learner practicing the modeled behavior; 4) feedback, which is provided by the instructor or trainer and serves both as a constructive tool to improve performance and as a social reinforcement for the acceptance of the new behavior; and 5) transfer of training (which the FAA calls transfer of learning) in which the learner applies the newly acquired behavior in context (Decker & Nathan, 1985). Flight instructor training outlines each of these steps as important components of an effective teaching process (FAA, 1999; Jeppesen, 2002).

Burton, et al. (1984) used learning to ski as an example of modeling and mastery learning. The beginning skier models the behavior of the instructor, who helps him set small, realistic goals as intermediate levels of expertise that he can master incrementally. Because the learner sees the instructor excel at skills he has great desire to master, and the instructor gives him the tools and support necessary to progress and master each step, he feels confident to progress to the next more difficult task (self-efficacy). The environment also plays a key role in ski instruction: for
example, progressing from more gentle to steeper slopes incrementally helps the skier build confidence each time he masters the new slope. Similarly, introduction of the graduated length method (initial use of shorter skis, then graduating to longer and longer skis as skills progress) takes advanced ski instruction to a completely new level using this mastery learning technique.

Learning to fly is much like learning to ski because it is a complex skill in which the starting and final states are far apart. Additionally, the consequences of not mastering flying skills at many levels can be deadly from beginning to advanced skills such as instrument flying or aerobatics. Learners may become doubtful or even fearful if their belief in their ability to master skills is compromised. For this reason, the flight instructor must become a trusted mentor who builds and supports the learner's sense of self-efficacy and facilitates mastery learning. The flight instructor facilitates mastery by separating complex skills into simpler subskills that can be mastered sequentially and progressively. To do this, the instructor models a task (subskill) that the learner can perform successfully and helps the learner set the goal of mastering each subskill in turn. This process reinforces the learner's belief in their ability to perform tasks as they continue to master each subskill. When the learner has mastered a number of subskills, those skills can be fitted together as components of the more complex skills required in advanced maneuvers such as takeoff and landing, instrument navigation, or barrel rolls. Using modeling and mastery learning, the student pilot gains confidence in her ability; she also gains insight into how to perform a skill, and when to combine skills in various flight operations.

A type of training widely used in the aviation industry is known as line oriented flight training (LOFT) and is modeled after the original training implemented by Northwest Airlines in the late 1970's (Kern, 1998). In this type of training, students fly a complete flight profile, including departure, enroute, arrival, and approach segments, in which specific training challenges have been predetermined and are introduced during the flight (e.g., systems and equipment malfunctions, traffic delays, weather hazards, diversions). This type of training allows for standardization and allows each student to receive the same challenges and flight scenarios regardless of
when or by whom the training is conducted. Current student training in most collegiate flight programs and many private schools makes use of LOFT training.

2.5.2 Summary

FAA approved flight training is solidly grounded in educational theory and standardized practice. Flight instructors are trained in these theories and practices, and it is important for any training program to adhere to standardized and proven methods and design. The CTM training course in this study followed common practice for design of FAA approved training curricula, and both design and delivery were thoroughly grounded in social learning theory (specifically strategies related to social cognitive learning theory as just described), constructivism, modeling, and mastery learning. Additionally, training was conducted by an experienced certified flight instructor. The objective was to remove as much ambiguity as possible regarding legitimacy of the training course as well as to use teaching and learning methods already familiar to pilots who participated in the study.

2.6 Literature Review Summary

This review has touched on a small portion of the vast body of literature regarding the development and evolution of aviation Human Factors, human error, and aeronautical decision making. It is ultimately unreasonable to expect pilots to operate completely without error. The crux of reducing human error is to enhance a pilot’s ability to recognize, analyze, and mitigate error before it adversely affects the safety of the flight. Pilots study and practice skills and concepts related to aeronautical decision making (ADM), but the degree to which ADM is integrated into flight training is highly variable. Although airlines and military training programs are structured to address ADM at all levels and both FAA and ICAO have recommended that ADM be
integrated into general aviation flight training, many programs lack proper or sufficient ADM training in their curricula.

The aspect of human error and ADM relevant to this study is that of a pilot’s multitasking ability, specifically concurrent task management. A pilot’s ability to perform multiple, simultaneous tasks is closely related to her cognitive abilities to handle multiple inputs and manage her attention effectively. Several cognitive processing models, including single channel, single resource, and multiple resource theory can be applied to better understand how a pilot will deal with incoming information, manage that information to prioritize tasks, and select when to shed tasks. It is likely that pilot ability to effectively process inputs and prioritize and execute tasks includes factors that vary greatly from one individual to another. However, a certain level of multitasking ability is required in order to perform even at a basic level adequate to achieve pilot certification, so that pilots in general comprise a group that has been pre-selected to some extent based on their demonstrated ability to properly prioritize and execute tasks in the context of flight operations. In order to increase pilot multitasking ability it seems reasonable that a system could be designed to help pilots better prioritize tasks in the cockpit.

To define such a system Funk et. al. (2003) laid a solid foundation by identifying factors which affect task prioritization. Because the use of technological aiding, such as the cockpit task management system and the Agenda Manager (Funk et. al., 2003) may actually increase pilot workload, the next logical step was to ask the question as to whether pilots can be specifically trained to better prioritize tasks. The aviation training industry in fact utilizes several methods such as Standard operating procedures, checklists, mnemonic memory aids, and cockpit flow checks to help pilots prioritize tasks. However, there is not currently an ADM training course specific to task prioritization in the cockpit.

Previous experimental studies conducted to address the question regarding whether or not prioritization training could be effective in reducing pilot error produced inconclusive results (Funk et. al., 2003), but those studies had several limitations, including instructional design and delivery, inadequate equipment, lack of
sample homogeneity, and small sample size. This study addressed those issues by utilizing FAA approved methods and practices for instructional design and delivery, FAA certified instructors, FAA approved FTDs, a more homogeneous sample groups, and a larger sample size (N = 27).

In general aviation, the vast majority of training programs introduce the aviate, navigate, communicate (ANC) hierarchy early in pilot training. Pilots are taught to use checklists, standard operating procedures, and flow checks in order to help them prioritize tasks. Additionally, they learn to constantly question their situation in order to stay ahead of the airplane and be able to act and react to inputs in the most effective manner consistent with flight safety. However, the large number of aviation accidents and incidents that can seemingly be attributed to misprioritization of tasks indicates pilots are not applying procedures and prioritization schemes as effectively as possible. Furthermore, common knowledge, including this investigator’s personal experience, indicates that pilots often misprioritize tasks, especially communication tasks, resulting in adverse consequences for flight efficiency and safety.

Some theoretical researchers used an aviate, navigate, communicate, manage systems (ANCS) scheme for prioritizing tasks (Funk et. al., 2003; Wickens et. al., 2003). However, general aviation flight training uses an ANC hierarchy (Chappell, 1998; FAA, 1999; Jeppesen, 2001, 2003a, 2003b; Kern, 1998; Kershner, 1998; Machado, 2001, 2003; Thom, 1991). This investigator believes the difference represents a gap between theory and practice and was more interested in practical application of the results, thus defined prioritization errors using the ANC scheme found in the flight training literature.

The CTM training course was grounded in learning theory relevant to flight training as accepted by the FAA with particular attention to theories of constructivism, mastery learning, and social cognitive learning in the hope that the training course might be more transferable to other flight training arenas if experimental analysis showed an improvement in task prioritization performance. The hope is that these experimental results will pose questions and provide recommendations to guide further research.
3 RESEARCH METHOD

This study was an experimental analysis of observed changes in task prioritization performance for a group of Central Washington University (CWU) flight technology students who participated in a CTM short course. Data gathered during a pilot study six months prior was used to establish procedures and define the method used to evaluate CTM errors. Pilot study criteria are discussed in context in this chapter and a concise description of the pilot study is presented in Appendix D. Participants comprised a relatively homogeneous group of pilots enrolled in the CWU accredited university flight technology program and were enrolled in the CWU flight training syllabus instrument pilot intermediate training stage or higher. Details of participants and research methods are described in the following sections.

3.1 Hypotheses

In addition to describing experimental results, data were used to test the following null hypothesis:

\[ H_0 \] – After completing a CTM specific training course, pilots will show no change in task prioritization errors.

If the null hypothesis could be rejected, the alternative hypothesis proposed was:

\[ H_a \] – After completing a CTM specific training course, pilots will show a change in task prioritization errors.
3.2 Experimental Design

A pretest-posttest control group design as defined by Campbell and Stanley (1963) was used because it provides good control for various elements of internal validity (Campbell & Stanley, 1963; Christensen, 2001; Gall, Gall, & Borg, 2003; Russ-Eft & Hoover, 2005). A Solomon four group design was considered, since it could also address possible errors related to a practice effect. However, the sample size (originally N = 35) was not large enough for statistically significant results using the Solomon four group design (Keeves, 1988; McClave & Sincich, 2003).

A repeated measures analysis (Christensen, 2001; Keeves, 1988) was also considered. Repeated measures is a within-subjects design and in such an experiment participants would have been observed at two or more points in time under two or more experimental conditions, such as in a pretest-posttest scenario. An advantage of using a within-subjects design stated by Christensen (2001) is...

...the fact that the investigator need not worry about creating equivalence in the research participants, because the same participants are involved in each treatment condition. In other words, participants serve as their own control." (p. 253)

However, the design has serious drawbacks, including interference between treatment conditions, or a sequencing effect in which participants' performance after any given treatment of the independent variable is influenced by previous treatments and is generally difficult to control (Christensen, 2001; Myers, 1987). Because this study involved only a single dependent variable (CTM error) and a single independent variable (the CTM training course) and because homogeneity of sample groups was controlled through random assignment and use of pilots with similar training and experience, the within-group repeated measures design was not considered the most appropriate design. Interestingly, results from this study as discussed in Chapter 5 suggest that a repeated measures design might be an appropriate follow-on technique to address questions that arose from data in this experiment.
The pretest-posttest control group design was chosen because it fit the available number of potential participants, the single independent variable design, the relative homogeneity of participants, and the minimization of practice effect due to all participants having previous flight time on the Frasca 141 flight training device (FTD).

After participants volunteered for the study and signed an informed consent document (Appendix A) they were each assigned a three digit ID number between 100 and 400 using the random number generator function in MS Excel Office 2003 software. Then each pilot ID number was randomly assigned to either the experimental or control group by flipping a coin. Random assignment minimized differences in pretest CTM performance between the experimental and control groups. Pilots were not told which group they were assigned to until after they completed the pretest flight. Consideration was given to using pretest error scores to create experimental and control groups matched on those pretest scores, but those scores were fairly randomly distributed, even for pilots of varying flight time, instrument time, and Frasca 141 time (see Figures 3.1 through 3.6 and related discussion). Based on those criteria which indicated the group comprised a relatively homogeneous sample that method of assignment did not seem warranted. After all pilots flew the pretest flight and scores were analyzed, each pilot was told into which group they had been selected.

Pilots in both groups flew pretest and posttest flights on the Frasca 141 but the treatment (CTM training course) was only administered to the experimental group. All participants in the study logged flight time in airplanes (between 1.5 – 3.5 hours) during the interim between pretest and posttest observations. Two participants logged additional FTD time, but not on the Frasca 141 FTD.

Both groups flew the same LOFT scenario for the pretest and posttest FTD flights. A question arose during experimental design as to whether use of the same LOFT for both the pretest and posttest was appropriate. Because participants in the experiments fly instrument approaches and procedures as part of their normal training it was not likely they would remember specifics of that particular LOFT two weeks later during the posttest. Interestingly, some of the participants commented that they
expected to fly a different LOFT scenario and thus had not given any further thought to the first LOFT. Additionally, data from the pilot study did not indicate any strong practice effect from using the same LOFT in both flights. The final reason the same LOFT scenario was used for both flights is that it was considered a greater challenge and a greater threat to internal validity to attempt to design a second scenario that was equivalent with respect to pilot workload and CTM challenges than it was to use the same LOFT for both pretest and posttest. It was not feasible to evaluate whether or not two different LOFT scenarios were equivalent in those regards.

3.3 Participants

A goal of this study was to use participants who comprised a relatively homogeneous group with respect to their flight experience, standardization of training, and previous time on the Frasca 141 FTD used for the experiments. Each pilot reported those data on a pilot questionnaire (Appendix A) just prior to their pretest flight. Although all the specified criteria were not identical for each participant, data indicate that the two groups were equivalent and that each participant’s variation in total logged flight time, stage of training, instrument time, time on all FAA approved FTDs, and time on the Frasca 141 did not correlate to their pretest CTM performance (Figures 3.1 – 3.5 and Table 4-1). Regression analysis (Table 3.1) correlated each pilot’s total flight time, stage of training, instrument time, and Frasca 141 time to pretest CTM error scores. Using the bivariate correlation coefficient interpretation of Newton and Rudestam (1999), correlations were none to weak for all those parameters. Regression showed low to moderate correlation between total FTD time and pretest CTM error scores (Table 3.1).
Figure 3.1. Graph showing relationship of total flight time and pretest CTM error scores for all participants. The highest possible number of errors = 20. Regression analysis yielded an $R^2 = 0.036$ for the control group and $R^2 = 0.148$ for the experimental group.

Figure 3.1 shows little to no relationship between each pilot’s total logged flight time and their pretest CTM error score. Indeed, the pilot with the highest logged flight time (400 hours reported) made the greatest number of CTM errors, while low time pilots made relatively low to average number of errors.
Figure 3.2. Graph showing relationship of stage of training and pretest CTM error scores for all participants. The highest possible error score = 20. Stages of training indicate that at the time of the study pilots were enrolled in the following portion of the CWU flight training syllabus: 1) instrument rating intermediate stage; 2) instrument rating final stage; 3) commercial pilot cross country stage; 4) commercial pilot advanced maneuvers stage; 5) commercial pilot final stage; 6) flight instructor initial stage; 7) flight instructor intermediate stage; 8) flight instructor final stage; 9) multiengine pilot initial stage; 10) multiengine intermediate stage. Regression analysis yielded an $R^2 = 0.018$ for the control group and $R^2 = 0.021$ for the experimental group.

Figure 3.2 shows little to no relationship between stage of training and pretest CTM errors. The majority of participants were enrolled in the first four stages of training, with pretest error scores varying between one and nine. The other six had pretest error scores varying from one to 11.
Figure 3.3. Graph showing relationship of total instrument time and pretest CTM error scores for all participants. The highest possible number of errors = 20. Regression analysis yielded an $R^2 = 0.002$ for the control group and $R^2 = 0.043$ for the experimental group.

Figure 3.4. Graph showing relationship of total FTD time and pretest CTM error scores for all participants. The highest possible number of errors = 20. Regression analysis yielded $R^2 = .001$ for the control group and $R^2 = .202$ for the experimental group.
Figure 3.5. Graph showing relationship of total Frasca 141 FTD time and pretest CTM error scores for all participants. The highest possible number of errors = 20. Regression analysis yielded an $R^2 = 0.097$ for the control group and $R^2 = 0.106$ for the experimental group.

Figures 3.3 through 3.5 show little to no relationship between pretest CTM errors and instrument time, total time logged on all FTDs, and time logged on the Frasca 141 FTD. An outlier in the experimental group (pilot E5 had logged almost three times the mean Frasca 141 time) did not significantly affect the distribution of those criteria. Regression analyses of the data with pilot E5’s scores removed show similar results.

Table 3.1 shows regression analysis of the criteria just discussed. As a descriptive statistic $R^2$ represents the amount of variability in one variable that can be explained by the other (George & Mallery, 2005; McClave & Sencich, 2003; Newton and Rudestam, 1999). For this data $R^2$ represents the variability in pretest error scores explained by each of the selected criteria (total flight time, instrument time, stage of training, FTD time, or Frasca 141 time). Based on the interpretation of Newton and Rudestam (1999), $R^2$ for each of the selected criteria show little to negligible relationship with pretest CTM error scores for either group of pilots.
Table 3.1. Relationship of reported pretest criteria and pretest CTM error scores according to group. The coefficient of determinations showed relationships were negligible to weak, indicating the two groups could be considered equivalent and relatively homogeneous with respect to pretest variables.

<table>
<thead>
<tr>
<th></th>
<th>Total time</th>
<th>Stage of training</th>
<th>Instrument time</th>
<th>FTD time</th>
<th>Frasca 141 time</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control group Pretest scores</td>
<td>$R^2 = 0.036$</td>
<td>$R^2 = 0.018$</td>
<td>$R^2 = 0.002$</td>
<td>$R^2 = 0.001$</td>
<td>$R^2 = 0.097$</td>
</tr>
<tr>
<td>Experimental Group pretest scores</td>
<td>$R^2 = 0.148$</td>
<td>$R^2 = 0.021$</td>
<td>$R^2 = 0.043$</td>
<td>$R^2 = 0.202$</td>
<td>$R^2 = 0.106$</td>
</tr>
</tbody>
</table>

Based on the data just presented, pilots in the two groups could be considered equivalent with respect to pretest criteria and CTM performance. There was little to no relationship between the variables of stage of training, total flight time, instrument time, time in FTDs, and time logged on the Frasca 141 and that of pretest CTM errors. This was confirmed by an independent samples t-test comparing pretest mean error scores for the two groups which gave $t = 0.14$ at $p = 0.88$. Therefore, large variations in posttest CTM error scores could more likely be attributed to the independent variable (training course) rather than to pretest variability between groups.
3.4 Independent Variable

The independent variable, or treatment condition, for this experiment was the CTM short course in which experimental group pilots participated during the interim period between the pretest and posttest FTD flights. During that same period control group participants did not engage in any additional activities related to the study. The discussions in Chapter 2 set the stage for the design and delivery of the CTM training course as well as the investigator's experience designing and teaching similar courses. The syllabus used for both the pilot study and the main study comprises Appendix B and details of specific learning activities are available on request. The course consisted of two learning sessions 10 days apart that included reading, self-study, cooperative learning activities, guided discussion, and a reflective homework assignment. Learning outcomes and assessment strategies were clearly stated at the beginning of the course and the investigator, who is a certified and experienced instrument flight instructor, conducted the training.

Because Funk et al. (2003) determined that procedure was the primary reason pilots prioritized tasks, which is corroborated by Chappell (1998) and Kern (1998), and because pilot training world-wide is based on procedural training and discipline, the training course emphasized procedural discipline with respect to task prioritization, including proper use of checklists, standard operating procedures, mnemonic aiding devices, situational awareness, and cockpit flow checks.

In order to integrate the concepts of mastery learning (Bloom, 1974; Burton, Brown, & Fischer, 1984) and scaffolded instruction (Bruning et. al. 1995) the course necessitated a high level of participation on the participants' part which was manifested through their identifying CTM errors from accident and incident data, proposing solutions to CTM errors, and reflecting on their personal flights conducted during the training period with regards to CTM. Additionally, the investigator actively encouraged participants to enhance their self efficacy and build on principles of social learning theories (Bandura, 1977; 1986; Schunk, 2003).
The first learning session consisted of a class discussion of selected materials related to aviation human factors, aeronautical decision making, situational awareness, workload management, and cockpit task management. Participants had prior knowledge of all those concepts from previous coursework and studies, thus the course did not introduce any new concepts but rather emphasized task prioritization as an important element of human factors. Participants analyzed accident and incidents taken from the NTSB and NASA with respect to CTM errors and participated in class discussions of those data.

During the time between sessions participants were asked to reflect on at least one of their normal flights with respect to CTM concepts and how they perceived their awareness of it to influence their in-flight decision making. Some students reflected in writing and some during a verbal debriefing with the investigator.

The second class session included an activity in which participants acted out role-playing scenarios designed to give insight into their reactions and behavior in the cockpit when confronted with CTM challenges. They also participated in a class discussion of strategies to improve pilot task prioritization performance and a guided discussion of the outcomes. A short quiz was given at the end of the second session to evaluate each pilot’s progress and identify areas of improvement.

Pilots in the control group had the opportunity to participate in an identical CTM training short course after the experiment was over to reduce any possible adverse effects of their not receiving training should the course prove beneficial.

3.5 Dependent Variable

The dependent variable was defined as the performance (task prioritization error rate) for pilots in both the experimental and control groups. To establish the parameters that defined the dependent variable, the task prioritization challenges and errors were defined based in part on CTM theory developed by Funk, et al., (2003), but specific challenges and errors were based on the aviate, navigate, communicate
(ANC) hierarchy (Chappell, 1998; FAA, 1999; Jeppesen, 2001, 2003a, 2003b; Kern, 1998; Kershner, 1998; Machado, 2001, 2003; Thom, 1991) as well as on procedures, checklists, and flow checks used in standard FAA flight training. Each task was defined as follows:

- **Aviate task:** Included all items related to aircraft operation: airspeed, altitude, climb or descent rate, lift, thrust, and drag; e.g. primary aircraft control inputs (pitch, power, yaw, and roll), operation of lift and drag devices (flaps) and operation of primary engine systems.
- **Navigate task:** Included items related to the current and future position of the aircraft, including vectors, course intercepting and tracking, identification of intersections and waypoints, and programming and operating the GPS and other navigation radios.
- **Communicate task:** Included communications with ATC.

This study did not use the ANCS task prioritization hierarchy from previous studies by Funk et al. (2003) for reasons related to standardization of training and both internal and external validity as discussed in the Chapter 2 and the following sections.

3.6 Internal Validity

In an experimental study there is always the question of whether the observed results are really a function of the treatment being tested, which depends greatly on the experiment’s ability to control extraneous variables that might affect the outcomes. As discussed in Chapter 2, previous CTM training studies conducted by Bishara (2002) and Funk et. al. (2003) did not control for several extraneous variables, which may have contributed to lack of internal validity as discussed by Funk et al. (2003). Unless some acceptable measure of control is gained over extraneous variables any statement based on experimental results may not be valid. In order to increase internal validity for this study, criteria identified by Funk et al. (2003) were addressed in
addition to criteria that emerged during the pilot study and during experimental design. Controlling some of those criteria was relatively straightforward, while control of others was more difficult when unforeseen events arose as discussed in the following sections. The criteria used here to evaluate internal validity are based primarily on those specified and discussed by Campbell and Stanley (1963) and Gall et. al. (2003).

3.6.1 Pretest Variability and Differential Selection

To control for pretest variability participants comprised a relatively homogeneous group with respect to standardization of training and previous experience. Also, participants were randomly assigned either to the experimental or control group and all had logged previous time on the Frasca 141 FTD. Because stage of training, previous flight time, instrument time, total FTD time, and Frasca 141 time varied for all participants those times were compared with pretest CTM error scores to determine the amount of pretest variability. There was no correlation between Frasca 141 time and pretest CTM error scores (Figures 3.1 – 3.5 and Table 3.1).

The criteria of participant homogeneity and random assignment also provided tighter control over possible effects of differential selection. Campbell and Stanley (1963) noted that “selection is ruled out as an explanation of the difference to the extent that randomization has assured group equality at time R [random assignment]” (p. 15). There was good control of both pretest variability and selection based on pretest data presented in Figures 3.1 through 3.5 and Table 3.1 in that the two groups were equivalent with respect to their CTM error scores, regardless of variability on several pretest criteria such as flight time, instrument time, stage of training, and Frasca 141 time.
3.6.2 Practice Effect

Practice effects were minimized partly by the condition that all pilots had previous time on the Frasca 141 FTD, so they should not have been learning how to operate FTD during observation flights. Also, because the same LOFT scenario was flown by all pilots for both the pretest and posttest flights there might have been a practice effect due to repetition of the same flight, but it is expected that the effect would be the same for pilots in both groups.

If there were a practice effect from the pretest flights it would be expected to show up in the control group performance data. Comparison of pretest and posttest error data from the control group did not indicate a practice effect due to repeating the same LOFT scenario for both flight trials. In fact, several pilots in the control group made more errors in the posttest flight than they did in the pretest, as shown later in Figure 4.1. Also, the specific errors and types of errors they made in the posttest were not the same as those in the preflight. For example, a pilot from the control group who made the same number of errors in the posttest as in the pretest did not in general make the same types of errors or the same errors. Therefore there did not seem to be a practice effect due to pretest sensitization.

It is conceivable that pilots with less instrument time or less time on the Frasca 141 might show more of a practice effect from the pretest since they would not be as familiar with the procedures or the FTD. However, data from Figure 3.1 through 3.5 and Table 3.1 indicate that variations in stage of training, total flight time, instrument time, and FTD and Frasca 141 time did not correlate with pretest CTM error scores, and such lack of correlation suggests there was no practice effect.
3.6.3 History Effect

In order to minimize history effects as discussed by Campbell and Stanley (1963) the time between pretest and posttest observational was relatively short (i.e., two weeks) so that any extraneous variables due to learning from other events besides the experimental treatment could be minimized. However, that short time period was considered a compromise, since it could have served to decrease external validity. It seems probable that any well designed training course would have some effect over the short term but might not have a lasting effect. Therefore, the time for the experiments was selected at three weeks to maximize the experiment’s ability to test beyond a short term learning effect but to minimize for history effects. Also, the investigator believed that any longer time between observation flights might contribute to mortality issues as defined by Campbell and Stanley (1963) which would also affect internal validity.

3.6.4 Experimenter Bias

In any study it is possible that expectations on the part of the investigator could be transmitted to participants or the people conducting and scoring the experiments. Two issues related to bias are inter-tester reliability, as defined by Gall et. al. (2003) and intercoder reliability as defined by Neuman (2000).

Inter-tester reliability is the bias that may be introduced by the person administering the test, the person scoring the test, or both. To increase inter-tester reliability different people should administer and score a test (Gall et. al. 2003). Using two instructors for each observation flight was problematic because flights were conducted on two FTDs simultaneously over a two day time period so that all data were collected at about the same time (observation flights were within 48 hours of one another). Because there was only a limited pool of qualified instructors there were not always four instructors available all the time during data collection. Therefore the
instructor administering the test (acting as ATC) also scored the test. To minimize potential bias the investigator personally briefed each instructor on the purpose of the study and the potential pitfalls associated with inter-tester bias. Additionally the instructors did not know whether pilots were in the control or experimental group.

Intercoder reliability relates to the consistency of scores obtained by multiple observers of the same phenomenon or test. Intercoder reliability is tested by having several coders measure the exact same thing, then comparing those measures (Neuman, 2000). To address issues related to intercoder reliability the investigator thoroughly trained each instructor regarding the definition of CTM errors and the specifics of the LOFT. Each of the ATC calls, procedures, and 20 CTM challenges were reviewed individually to assure that every instructor understood the criteria being used. Then the investigator demonstrated the entire LOFT using the Frasca 141 and a “dummy” participant while the instructor observed. Finally, the instructor conducted the LOFT using the dummy participant while the investigator observed and evaluated their ability to properly conduct of the LOFT and their definition of CTM errors, and any discrepancies were rectified prior to the pretest observations. Finally, scores from live observations were compared with a second scoring conducted by a different instructor who watched the videotape of the FTD flight. The assignment of instructors was based on their availability, and each pilot was scored by at least three different coders. None of the coders knew which participants were in which group. Coders took notes both during live observations and when viewing videotapes so the investigator could view and resolve discrepancies. In general there were not many discrepancies between coders but for a few that could not be resolved it was recorded as a non-error in the data.
3.6.5 Mortality

Mortality issues posed a threat during the experiment. Initially 35 participants signed up for the study, completed the informed consent documents, and were assigned a participant ID number and randomly assigned to either the experimental or control group. The control group originally had 18 and the experimental group had 17 participants. Before the pretest flight one participant withdrew due to illness. Another participant came for the pretest flight, filled out the pretest questionnaire, and went through the preflight procedures briefing. However, just before commencing the flight in the FTD that individual decided not to participate. The investigator personally assured that person that any information they provided for the experiments would be destroyed and shredded all associated documents in full view of the participant.

Another mortality issue arose during conduct of the experiment. A total of six instructors went through the training as just described and served as raters either during real time or during the second set of scoring of videotapes. One instructor conducted the first two experimental observation pretest flights correctly, but for some reason during the following four flights the instructor diverged from the LOFT script; ATC calls, vectors, and procedures did not follow the design. The discrepancy was not discovered until the second coder observed the videotape of those pretest flights. Therefore, those four participant’s data had to be deleted from the study. So as not to cause any more adverse effects the investigator did not disclose that information to the participants but conducted the posttest flights normally. That instructor was not used again in the study.

The final mortality issues occurred when one participant could not make it to the posttest flight due to illness, and one who discontinued the posttest flight 10 minutes before the end of the flight. That participant did not say why they quit, just that they did not want to continue. Those two participant’s data were also deleted from the study. A total of eight (five from the control group and three from the experimental group) were lost due to mortality, leaving a final N=27 for the analysis (13 in the control group and 14 in the experimental group).
3.6.6 Contamination

To minimize contamination due to interaction between the two groups of participants, instructors hired to run the FTD flights, experimental coders, and all participants were instructed not to discuss the experiments with anyone else during the entire time. The only exception was during the CTM course when elements of CTM were discussed. However, specifics of the experiments were not. Because all the individuals who participated in the study typically see one another on a daily basis, contamination could have potentially posed a large threat to internal validity. However, because of that same contact the investigator had many opportunities to remind participants not to discuss the experiments (and did so repeatedly). Absolute certainty with respect to lack of contamination is not possible, but the investigator believes that all participants and instructors understood the importance of not discussing their role or specifics of the experiment and that contamination was not a factor.

3.6.7 Defining and Recording CTM Errors

A major factor believed to increase internal validity was the definition of the CTM errors themselves. As discussed in Chapter 2, the investigator questioned whether some errors recorded by previous studies were valid in the context of current flight training practices and procedures. For example, Funk et al. (2003) recorded an error when a pilot experienced a magneto failure and did not report a turn to final approach within 10 seconds of the requested time. However, a magneto failure constitutes a situation than could lead to a loss of power and an emergency situation as per 14 CFR Part 91.205. Thus a magneto failure is directly related to operation of the aircraft (an aviate task) and would take precedence over other tasks, such as reporting the turn to final approach. Kern (1998) stated that “handling a critical action emergency procedure [which] is also part of the aviate step” (p. 91). A properly
trained pilot would aviate (control the aircraft) first, then use a standard flow check or procedure to troubleshoot the magneto failure and prevent possible engine stoppage, and navigate (make the turn to final), before making a report to ATC regarding their position, particularly when under positive radar control. Most procedures instruct the pilot to tell ATC to "stand by" until an urgent or emergency situation can be controlled. Finally, many of the challenges presented in Bishara's (2002) and Funk et. al.'s (2003) study were unrealistic, such as raising the landing gear when a pilot had already put it down, or causing the carburetor heat control to "pop on" which would not occur in a real aircraft. See Funk et al. (2003, Appendices C, D, and E) for a complete presentation of those challenges and definitions.

The performance definition of errors in previous studies was not consistent with criteria outlined by the FAA and with which pilots are familiar, which might have affected internal validity. A performance error in this study was defined according to the FAA Practical Test Standards stated by FAA-S-8081-4C with respect to altitude, airspeed, heading, intercepting and tracking course, use of checklists, procedures, and ATC communications. Four types of CTM errors were used, including aviate/navigate (A/N), navigate/communicate (N/C), aviate/communicate (A/C) and aviate/aviate (A/A) (see appendix D). The aviate/aviate error was defined as one that occurs when a pilot performs a lower priority aviate task with respect to procedure or status over a higher priority one, for example deviating from altitude or airspeed while conducting a systems flow check or attending to a malfunction that is not directly affecting the flight parameters of the aircraft.

Previous studies also did not use standardized procedures which might have introduced an additional unknown parameter that the pilot had to contend with, possibly influencing internal validity. For example, Bishara (2002) stated "All communications frequencies on the IFR charts and approach plates were disregarded" (p. 62). Additionally, "The pilot was instructed to disregard the SIDS (Standard Instrument Departures) and the STARS (Standard Terminal Arrival Routes) for both airports in each scenarios" (p. 62). Instead, Bishara (2002) created a route of flight from a Class D airport to an uncontrolled airport within Class E airspace but had the
pilot pretend to be in positive radar control at all times and simulated the level of control as Class C or Class B radar service.

There was no explanation in Bishara’s (2002) study as to why he chose to change the FAA published procedures or to deviate from standardized procedures. Based on several thousand hours of experience as a flight instructor, the investigator believes that such a deviation could have caused confusion as well as abnormally increasing the pilot’s workload since they could not refer to standard charts for frequencies and airspace requirements as they would be accustomed.

Standard FAA instrument training includes a required “approach briefing” in which the pilot verbally calls out all frequencies, procedures, altitudes, course information, weather minimums, and any special instructions based on the standard instrument approach chart prior to commencing the approach. To control for those types of threats to internal validity in previous studies the LOFT for this experiment used standard FAA flight profiles, departures, arrivals, approaches, and communication and navigation frequencies, and Instrument approach procedure (IAP) charts. The LOFT is described in detail in Appendix C.

Funk et. al. (2003) used specific CTM challenge points that were pre-determined and embedded within their simulated flight scenarios. Because errors occur throughout the flight, and not just at specified points, the question arose as to whether it might be more appropriate to record all errors observed instead of at specific predetermined points within the LOFT scenario. The pilot study was designed to investigate that question using a LOFT scenario that placed pilots in a high workload environment for most of the flight so that the possibility for CTM errors was maximized, and errors were recorded and classified as they were observed. The problem that arose from using that technique was that it was extremely time consuming because the errors had to be evaluated and defined individually. Additionally, the definition of errors required detailed knowledge of the exact procedures that were supposed to be used, both according to FAA regulations as well as the operations specifications of the aircraft being used, which would be difficult to replicate in other experiments. For those reasons the pilot study validated the
technique used by Funk et al. (2003) of using specific challenge points embedded within the flights as a way to gather data more easily and that data could be more easily transferred to other studies.

Based on the information gathered in the pilot study, as well as discussion with Ken Funk (personal communication, September 20 and October 13, 2004) the investigator decided to follow the example used by Funk et al. (2003) and define specific CTM challenge points within the LOFT scenario. By specifying 20 specific challenges, each pilot had an equal chance of making an error at each challenge point. The points at which pilots made the most CTM errors during the pilot study, as well as the types of errors they made, was used to define the position and kind of challenge point for the LOFT used in the experiments (Appendix D).

The LOFT placed the pilot in a relatively high workload environment and some challenges were embedded in the procedures themselves and did not need to be specifically introduced. For example, the ILS 13R approach into Boeing Field involves several unusual procedures with respect to pilot action if glideslope is inoperative, as well as a fairly complicated MAP that requires pilots to track the localizer course while identifying cross radials to initiate a climb sequence and a turn to the missed approach hold point (Appendix C). Thus the potential for error was inherent in the procedure itself and the challenge point only involved observing if the pilot prioritized tasks properly.

Other challenge points involved ATC calling the pilot just as he/she was executing some other important aviate or navigate task or failing a navigational aid or component during the approach or MAP. There were a total of 14 challenge points that presented 20 potential errors, including a possible of seven A/N errors, seven A/C errors, five N/C errors, and one A/A error that involved the formation of carburetor ice (challenge point 12, Appendix D).
3.6.8 Internal Versus External Validity

It would not be possible to definitively control for all threats to internal validity, but every attempt was made during the design and conduct of the experiments to minimize certain effects as just described. One problem that such a design faces is that the same criteria used to increase internal validity may decrease external validity. For example, the duration between flights (two weeks) was short enough to increase internal validity because it minimized possible history effects from extraneous learning, but it could also have made the results less meaningful with regards to the long term effects of training. If the study were able to comment on longer term training effects it could increase external validity.

Selection of the participants was another compromise between internal and external validity. Participants were chosen because they comprised a relatively homogeneous group with respect to previous training and experience. They were all enrolled in the CWU flight technology program standardized course of training. Homogeneity also means that results may only be generalized to very similar groups within the overall pilot population, e.g. students enrolled in other highly standardized collegiate flight programs at a similar level of training and proficiency. Other issues related to external validity are discussed in section 3.7.

3.6.9 Summary

Random assignment was used to help control for differential selection and possible pretest variation. Also, pretest error data was analyzed to determine a possible correlation with factors such as total flight time, instrument time, stage of training, total FTD time, and Frasca 141 time. No correlation was found, so participants were not paired for the posttest observations.

The experiment was designed and executed using input from previous studies by Funk et. al. (2003) and Bishara (2002). Specifics used to define the CTM errors
were based on actual errors made in the pilot study (Appendix D). The ANC prioritization hierarchy was used in order to be consistent with FAA and general aviation flight training practices. Performance criteria used to define errors was based on FAA-S-8081-4C with respect to altitude, airspeed, heading, intercepting and tracking course, use of checklists, procedures, and ATC communications.

Threats to validity that emerged during data collection, such as mortality, had to be dealt with in a way that minimized impact, and the investigator chose to delete all data from participants who dropped out of the study or whose data were compromised by incorrect experimental procedures.

Inter-rater and intercoder effects were minimized to the best of the investigator’s ability through extensive standardization of instructors who administered and scored the observations, as well having a second instructor code a videotape of the same flight. The training course followed industry and FAA standard procedure and was grounded in appropriate learning theory to enhance external validity.

3.7 External Validity

External validity reflects how well the results of an experiment can be used to project to a larger group or similar population beyond the group that was studied in the experiment. Several factors could affect external validity. In previous CTM studies external validity was compromised through the use of a non-standard and non-FAA-approved flight simulation platform (Microsoft Flight Simulator), which means that results might only apply to pilots, or non-pilots, who were familiar with operation of MS Flight Simulator. External validity for this study was improved by using an FAA-approved FTD that meets FAA requirements for logging instrument time. Specifics of the Frasca 141 FTD are presented in the next section.

Another factor that might affect external validity is the pretest sensitization effect, which means that the pretest could interact with the experimental treatment and
affect the research results. Pretest sensitization was also an issue in previous studies, and Funk et al. (2003) concluded there may have been a pretest sensitization that affected the posttest results; pilots may have still been learning how to operate the MS Flight Simulator software during all of the experimental flights.

The two-group experimental design used in this study did not control for sensitization as well as a Solomon four-group design would have, but the sample size (initially N=35) was not deemed large enough to use a Solomon four-group. However, pretest sensitization was addressed by selecting participants from a relatively homogeneous group of pilots with similar training and who had previous experience with LOFT.

The CTM training course was designed and taught in the same manner as other FAA flight training courses so that it would better represent standardized training. However, it could be argued that the results only represent the training as conducted by the individual teacher and there could be significant variability between teachers using the same syllabus and training materials. That is certainly a critical factor inherent to any such study, and although every attempt was made to conduct the training in as standardized a manner as possible, the only way to determine if the study evaluated the teacher would be to repeat the experiment using other teachers.

The extent to which results from this study can be projected to a larger group also depends on how representative the participants were of the larger pilot population. Some validity is therefore inherent based on criteria already in existence in the flight training industry. For example to become certificated, pilots must exhibit satisfactory performance relative to a highly standardized and detailed set of performance criteria specified in various FAA practical test standards (PTS) for each license or rating.

Because procedures specified by the FAA under 14CFR Parts 61, 91, and 141 are the same for all flight training and flight operations pilots are a uniquely standardized group. Pilots at the same stage of training should be roughly equivalent regardless of where they receive training. However, differences exist, especially between professional and collegiate flight training programs usually conducted under 14CFR Part 141, versus private flight training courses usually conducted under 14CFR


Part 61. The intent was for this study to be able to comment on results relative to students at CWU as well and to the possibility or relating those results to other collegiate programs or similar programs conducted under 14CFR Part 141.

3.8 Experimental Platform

Two identical Frasca 141 FTDs were used to collect data for this study (Figures 4.7 and 4.8). The Frasca 141s are fully approved for instrument training under 14 CFR Part 91 and Part 141. The Frasca 141 was chosen because it is the same FTD previously flown by all potential participants in their normal training activities. The Frasca 242 FTD used in the pilot study (see Appendix D) was considered, but only about a third of the potential participants had logged time on that FTD. The Frasca 141 was also chosen because the two could be operated simultaneously and data could be collected in a more timely manner.

The Frasca 141 can be programmed to simulate performance and systems of multiple makes and models of single engine aircraft. For this study the Frasca 141 FTDs were configured as Piper PA28R-200 aircraft since that is what participants had flown in both airplanes and FTDs during prior training. A full description of the specifics of flight controls, systems, avionics, GPS, and operating speeds and procedures for the FTD is available upon request. Briefly, the FTD was configured as a normally aspirated single engine fixed gear low wing aircraft with single boost pump and vacuum pump. The Avionics package includes a Bendix/King stack with dual KY 196 Coms, dual KN 53 nav radios, KDI 572 DME, KR 87 ADF, KT76A transponder, KMA 24 audio panel with marker beacons, and GNS430 IFR enroute and approach certified GPS. Figures 3.6 and 3.7 show the cockpit layout and instructor station of the Frasca 141 FTDs used in the study.
Figure 3.6. Photo showing wide view of the Frasca 141 cockpit. A full aircraft description is available on request.

Figure 3.7. Photo showing instructor station for the Frasca 141. The instructor can induce challenges via various environmental, systems, and performance parameters, including instrument and systems failures.
3.9 Data Collection

Approximately 75 CWU flight technology students met the eligibility requirements for the study and they were recruited through announcements in classrooms and flyers posted on bulletin boards in the CWU flight technology building and at the airport flight school facility. It is estimated that the majority of those eligible students were made aware of the study through those recruitment methods.

Potential participants were able to ask questions of the investigator at any time during the three weeks of recruitment and were briefed on specifics of the study. Each potential participant reviewed the informed consent document (Appendix A) verbally with the investigator and decided whether or not they wished to participate. A total of 35 pilots signed up and entered their names and contact information on a schedule sheet posted on the investigator’s office door indicating their preferred flight times over a two day data collection period for the pretest FTD flights. The control group initially comprised 17 participants and experimental group 16 participants. Due to mortality issues described in section 3.6.5 the final data included 14 in the experimental group and 13 in the control group.

For both pretest and posttest observations each participant in each group flew a one-hour simulated flight on the CWU Frasca 141 FTD. Pretest FTD flights were all scheduled over a single weekend and posttest flights over another weekend two weeks later. To avoid between-group interaction and possible contamination participants were asked not to discuss the experiment or the flights with anyone else. Instructors and coders were also asked not to discuss the experiments until after the study was completed.

The investigator called each participant the day before their assigned pretest flights to remind them of their assigned time. When they arrived for their assigned flight time participants completed the pretest questionnaire (Appendix A) and the investigator briefed them regarding what IFR approaches and procedures the flight would entail. The investigator provided all IFR charts, kneeboard, note paper, yoke clip, and flight timer. If participants wanted to use their own equipment that was
allowed, but they all had to use the IFR approach plates provided by the investigator. Each pilot had up to 15 minutes to brief the approaches and route of flight before starting their flight.

After their preflight briefing the investigator took each pilot into the simulator lab and assigned them to one of the two Frasca 141 FTDs. Because the instructors hired and trained to act as ATC and coders also teach in the CWU flight technology program, the investigator asked if they had flown with or given flight instruction to each participant. If so, that participant did not fly with that instructor. Every participant flew with an instructor who was not familiar with their previous flight performance.

Participants operated the Frasca 141 as they do in their normal instrument training and all assigned flight tasks followed standard operating procedures as designated by the Federal Aviation Administration (FAA) and the CWU standard course of training.

In order to conduct more flights during the two day period the FTD was left running so that pilots did not have to go through an engine startup and systems check on the ground. The FTD sessions all started with the aircraft positioned at the runway hold line and pilots were instructed to get IFR clearance and ATIS from ground control as they normally would before departure. Flights were scheduled every one hour and 15 minutes beginning at 0800 through 1800 each day.

After completion of the pretest flight participants learned which group they had been selected into and signed up for their posttest flight appointment two weeks later. Experimental group participants participated in a two week CTM short course as described in Appendix B. The investigator called each participant the day before their posttest flight to remind them of their assigned time.

To obtain data on CTM errors the instructors/coders referred to the scripted LOFT (Appendix C) and if they determined a CTM error was committed at each of 14 specific challenge points they circled the error and wrote a note as to specifics of the situation. If they were unsure they put a question mark by the error and wrote why. The second coder viewed the flight on videotape and did the same.
All flights were videotaped using VHS camera equipment mounted on a tripod to the left of the pilot. By setting up the shot through the side window of the FTD only one camera was needed to get a wide angle view of the whole instrument panel, both of the pilot’s hands, the yoke and engine controls, laptop or kneeboard, rudder pedals, and avionics. The pilot’s face was not visible in the tapes. The audio on the tapes was good and the investigator started each tape just before the pilot obtained takeoff clearance and spoke the three digit ID number into the camera. All ATC and pilot radio calls and pilots’ verbal callouts of flow checks, checklists, and approach briefings were clearly audible on the tapes. The investigator also viewed all the tapes after they were coded to observe specific examples of errors and to resolve any discrepancies between coders.

The definition of whether or not a CTM error occurred was based on the CTM LOFT in Appendix C, which was discussed and approved by the dissertation committee before the experiments were conducted. Specific aircraft performance criteria used to determine errors adhered to FAA Practical Test Standards stated by FAA-S-8081-4C with respect to altitude, airspeed, heading, intercepting and tracking course, use of checklists, procedures, and ATC communications.

A total of 20 possible CTM errors were embedded at 14 challenge points throughout the flight, and consisted of seven possible A/N errors, seven A/C, five N/C and one A/A error so data could be analyzed to investigate total errors as well as the relationship of specific types of prioritization errors relative to the treatment.

A separate hard copy of the LOFT scenario was used to record data for each pilot’s pretest and posttest flights. Once total error and component scores were compiled data was entered into both excel and SPSS software for compiling graphs, tables, and running statistical tests. Data were entered as a frequency distribution of total errors, and also broken down into specific types of errors for each pilot. Those data are presented in Chapter 4.
3.10 Data Analysis

Raw data were analyzed from frequency distributions of error scores and graphically to comment on relationships between CTM errors, types of errors, and changes in the experimental and control group pretest and posttest performance. Statistical analysis of the data was also used to test the null hypothesis.

Many research texts describe the null hypothesis significance test (NHST) as a standard or preferred statistical method (Christensen, 2001; Gall et. al., 2003; Kendrick, 2000). However, other authors caution use of NHST without verification from some other method (Cohen, 1994; Giles, 2000; Leong & Austin, 1996) and state that its use is controversial and surrounded by confusion (Cohen, 1994; Giles, 2000). Giles (2000) stated that NHST has caused so much confusion that some psychologists contend that it should be abandoned in favor of alternative methods. Cohen (1994) argued that if the null hypothesis really meant there were no difference whatsoever then it was certain to be rejected and therefore NHST was meaningless. He suggested researchers really meant the null hypothesis test to mean there is no meaningful difference, and suggested calling it the nil hypothesis. For this study the null hypothesis test represented one technique that was used in conjunction with other methods such as graphical and comparative analysis of data. Results of those analyses are presented in Chapter 4 and discussed in Chapter 5.

3.10.1 Statistics

Any statistic used to analyze data or to test a hypothesis should be chosen carefully and thoughtfully with respect to its appropriateness and accuracy. The decision whether to use parametric or nonparametric tests for significance should be based on the distribution with the objective of analyzing the data in as simple and meaningful manner as possible. Most importantly, results from statistical tests should be validated by other methods of analysis, such as scrutinizing the raw data.
When using parametric tests for descriptive statistics such as an independent samples t-test, it is assumed that samples exhibit homoscedasticity (homogeneity of variances), are normally distributed, and that the two samples are independent from each other. If the assumptions of normal distribution and homoscedasticity are not valid, then the chance of a Type I error in rejecting the null hypothesis is increased. For small samples, or when the assumptions cannot be met, non-parametric statistics may be more appropriate and more powerful. Non-parametric tests make no assumptions about the frequency distributions of the variables being assessed but they do require that the distributions have the same form.

Statistics literature is not specific with regards to what constitutes a large sample, since it depends on the context and distribution of each sample or population. However, samples with more than 30 degrees of freedom can generally be considered large enough to approach a normal distribution as defined by the “law of large numbers” (Christensen, 2001; Kendrick, 2000; McClave & Sincich, 2003).

Because the sample size in this study was less than 30 (N=27), tests for homoscedasticity and normal distribution needed to be conducted to see if parametric or nonparametric statistics were more appropriate. Additionally, the type of statistic most appropriate needed to be determined. Based on the type of experimental design used for the study, the following criteria were determined regarding what statistic would be used.

If assumptions for using parametric tests were not violated, the most direct and simplest parametric test for data from only two sample groups and one independent variable would be the independent samples t-test using the difference in mean error scores from each group. According to Campbell and Stanley (1963) the independent samples t-test is the most widely used and acceptable statistic for the pretest-posttest control group design.

If the samples did not exhibit all of the criteria necessary for using parametric tests then the non-parametric test chosen as most appropriate was the Wilcoxon Rank-Sum (Mann-Whitney) test as described in McClave and Sincich (2003). The Mann-Whitney is one of the most powerful nonparametric tests and only requires that
samples be independent and random (McClave & Sincich, 2003: Norusis, 1999). The test entails combining the two samples, sorting the result, and assigning ranks to the sorted values, giving the average rank between tied observations. If the two populations have the same distribution then the sum of the ranks of the first sample and those in the second sample should be close to the same value so it is similar to a t-test for testing the null hypothesis.

3.10.2 Tests for Homoscedasticity and Normal Distribution

In order to determine if necessary criteria were met to use parametric statistics as just discussed in section 3.10.1 an F-test as described by McClave and Sincich (2003) was used to test for homoscedasticity, or equality of variances between the two samples being compared. The test either accepts or rejects a null hypothesis that the two samples have equivalent variance based on comparison of the calculated value of F to a critical value of F. To do that the pretest-posttest difference in error score was calculated for each set of scores (control and experimental groups) and the F-test conducted using Microsoft Excel 2003 statistics F-test function. The test yielded an F = 1.34 at P = 0.029, which did not exceed the critical F value of 1.74, thus indicating the null hypothesis was not rejected and two samples exhibited homoscedasticity.

To test for normal distribution a Kolmogorov-Smirnov Goodness-of-fit test, or K-S test, (Chakravart, Laha, & Roy, 1967) was used. The test either accepts or rejects a null hypothesis that the data follow a specified distribution (in this case the normal distribution). To do that the test uses the mean and variance to compare the sample cumulative distribution to the hypothesized cumulative distribution. An advantage of the K-S test over other goodness-of-fit tests, such as the \( \chi^2 \) test, is that it uses exact values rather than approximations. Disadvantages are that the distribution must be continuous and the test is more sensitive near the center than at the tails (George & Mallery, 2005).
The K-S test is an empirical function based on the theoretical cumulative distribution for the distribution being tested. The hypothesis regarding the distributional form is rejected if the test statistic is greater than the critical value at a given level of significance. The K-S function in SPSS 11.0 was used to calculate the K-S statistic which yielded a $Z = 0.70$ at $p = 0.71$, which indicated the distribution could be considered normal for use in statistical analysis (George & Mallery, 2005). A Q-Q plot of the data performed in SPSS 11.0 also showed that a normal distribution was approximated. Together those tests indicated that parametric statistics could be used.

However, a histogram of the data (Figure 3.8) shows the distribution had relatively high amplitude with increased scores at both tails. Since a t-test is based on comparing means and is thus more sensitive to deviations at the tails of the distribution the investigator decided to use a more conservative nonparametric statistic and the Mann-Whitney test was chosen.

![Figure 3.8](image.png)

Figure 3.8. Histogram showing distribution of pretest-posttest difference in error scores for both experimental and control groups. Since the higher scores at both tails might have greater influence on a t-test, the nonparametric test was chosen.
3.11 Summary

Many criteria were used during experimental design to increase both internal and external validity. A relatively homogeneous group of pilots from a single standardized training program comprised two groups, an experimental and control group. Practice effects were minimized by using pilots already familiar with the FTD and the procedures used in the study, and pretest-posttest differences in control group error scores indicated there was no practice effect. The two groups were equivalent based on variables of previous flight time, instrument time, FTD time, and stage of training compared to their pretest flight CTM performance.

Training course design and delivery adhered to industry standards and was grounded in appropriate learning theory. Multiple measures were used to control for pretest variability, selection, maturation, history, and experimenter bias, including random assignment, use of multiple evaluators in real and recorded time, and timing of the experimental flights. Mortality was harder to control and the experiment lost eight participants due to various reasons. However, the attrition was similar for each group; the control group lost five and the experimental group lost three participants. The final distribution was 13 for the control group and 14 for the experimental group. Thus the groups were still equivalent based on an odd number of total participants.

The LOFT scenarios were designed and CTM challenge points defined based on observations and criteria established in the pilot study. Definitions of CTM errors adhered to FAA PTS with regards to performance standards.

Data were analyzed to determine whether parametric or nonparametric statistics were more appropriate. Although samples were random and independent, and the data exhibited homoscedasticity and approximated a normal distribution, because the data exhibited deviations at both extremes and had suffered some attrition during the experiment the Mann-Whitney test was chosen to test the null hypothesis.
4 RESULTS

In the first section raw data scores are presented in tabular and graphical format. Comparisons of various pretest-posttest errors both between and within groups are presented, including a breakdown into the three main error categories. In the second section the calculation of the Mann-Whitney test of the null hypothesis is presented.

4.1 CTM Error Data

CTM error scores are shown in Table 4.1 as pretest or posttest for each participant. Scores represent the number of errors out of a possible 20. The 20 possible errors were embedded in 14 challenge points throughout the LOFT scenario (Appendix C). Each participant’s original three digit ID number was recoded to add another level of protection for anonymity. Those codes (Table 4.1) distinguish pilots in the control and experimental groups (C1, E2 etc.). Table 4.1 also shows what each pilot stated on the pretest questionnaire with regards to their stage of training, total flight time, instrument time, time in all FTDs, and time in the Frasca 141 FTD prior to the pretest FTD flight. Data collected before the posttest flight revealed that all pilots had flown in airplanes between the pretest and posttest for an average of 2.3 hours, but none had flown in the Frasca 141 FTD. Three pilots in the control group and one in the experimental group flew an average of 3.6 hours in other FTDs.
Table 4.1. Frequency distribution of participant data and CTM error scores. Personal data was reported by participants just prior to the pretest FTD flight. Error scores represent the total raw CTM errors made out of a total of 20 possible.

<table>
<thead>
<tr>
<th>ID</th>
<th>Stage of Flight</th>
<th>Flight instrument</th>
<th>FTD</th>
<th>Frasca 141</th>
<th>pretest</th>
<th>posttest</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>number</td>
<td>training hours</td>
<td>hours</td>
<td>hours</td>
<td>hours</td>
<td>error</td>
</tr>
<tr>
<td>C1</td>
<td>1</td>
<td>95</td>
<td>25</td>
<td>5</td>
<td>5</td>
<td>2</td>
</tr>
<tr>
<td>C2</td>
<td>4</td>
<td>235</td>
<td>78</td>
<td>22</td>
<td>22</td>
<td>7</td>
</tr>
<tr>
<td>C3</td>
<td>2</td>
<td>99</td>
<td>27</td>
<td>10</td>
<td>10</td>
<td>3</td>
</tr>
<tr>
<td>C4</td>
<td>3</td>
<td>145</td>
<td>55</td>
<td>15</td>
<td>15</td>
<td>6</td>
</tr>
<tr>
<td>C5</td>
<td>3</td>
<td>200</td>
<td>50</td>
<td>15</td>
<td>15</td>
<td>1</td>
</tr>
<tr>
<td>C6</td>
<td>10</td>
<td>335</td>
<td>119</td>
<td>29</td>
<td>11</td>
<td>2</td>
</tr>
<tr>
<td>C7</td>
<td>4</td>
<td>260</td>
<td>76</td>
<td>15</td>
<td>15</td>
<td>1</td>
</tr>
<tr>
<td>C8</td>
<td>1</td>
<td>75</td>
<td>25</td>
<td>10</td>
<td>10</td>
<td>7</td>
</tr>
<tr>
<td>C9</td>
<td>1</td>
<td>115</td>
<td>40</td>
<td>12</td>
<td>12</td>
<td>4</td>
</tr>
<tr>
<td>C10</td>
<td>7</td>
<td>400</td>
<td>70</td>
<td>8</td>
<td>8</td>
<td>11</td>
</tr>
<tr>
<td>C11</td>
<td>1</td>
<td>112</td>
<td>26</td>
<td>15</td>
<td>8</td>
<td>3</td>
</tr>
<tr>
<td>C12</td>
<td>7</td>
<td>260</td>
<td>52</td>
<td>15</td>
<td>15</td>
<td>4</td>
</tr>
<tr>
<td>C13</td>
<td>4</td>
<td>208</td>
<td>57</td>
<td>27</td>
<td>27</td>
<td>9</td>
</tr>
<tr>
<td>Total</td>
<td>NA</td>
<td>2537</td>
<td>700</td>
<td>207</td>
<td>173</td>
<td>60</td>
</tr>
</tbody>
</table>

Control Group

<table>
<thead>
<tr>
<th>ID</th>
<th>Stage of Flight</th>
<th>Flight instrument</th>
<th>FTD</th>
<th>Frasca 141</th>
<th>pretest</th>
<th>posttest</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>number</td>
<td>hours</td>
<td>hours</td>
<td>hours</td>
<td>hours</td>
<td>error</td>
</tr>
<tr>
<td>E1</td>
<td>1</td>
<td>82</td>
<td>24</td>
<td>9</td>
<td>9</td>
<td>6</td>
</tr>
<tr>
<td>E2</td>
<td>1</td>
<td>100</td>
<td>26</td>
<td>6</td>
<td>6</td>
<td>2</td>
</tr>
<tr>
<td>E3</td>
<td>3</td>
<td>158</td>
<td>80</td>
<td>31</td>
<td>17</td>
<td>4</td>
</tr>
<tr>
<td>E4</td>
<td>7</td>
<td>259</td>
<td>47</td>
<td>31</td>
<td>17</td>
<td>1</td>
</tr>
<tr>
<td>E5</td>
<td>2</td>
<td>170</td>
<td>100</td>
<td>42</td>
<td>42</td>
<td>2</td>
</tr>
<tr>
<td>E6</td>
<td>1</td>
<td>150</td>
<td>50</td>
<td>20</td>
<td>20</td>
<td>6</td>
</tr>
<tr>
<td>E7</td>
<td>7</td>
<td>320</td>
<td>100</td>
<td>30</td>
<td>16</td>
<td>3</td>
</tr>
<tr>
<td>E8</td>
<td>6</td>
<td>260</td>
<td>40</td>
<td>15</td>
<td>11</td>
<td>7</td>
</tr>
<tr>
<td>E9</td>
<td>3</td>
<td>190</td>
<td>60</td>
<td>13</td>
<td>13</td>
<td>9</td>
</tr>
<tr>
<td>E10</td>
<td>3</td>
<td>188</td>
<td>55</td>
<td>18</td>
<td>18</td>
<td>2</td>
</tr>
<tr>
<td>E11</td>
<td>1</td>
<td>109</td>
<td>29</td>
<td>10</td>
<td>7</td>
<td>5</td>
</tr>
<tr>
<td>E12</td>
<td>3</td>
<td>146</td>
<td>50</td>
<td>12</td>
<td>12</td>
<td>9</td>
</tr>
<tr>
<td>E13</td>
<td>4</td>
<td>202</td>
<td>64</td>
<td>10</td>
<td>10</td>
<td>3</td>
</tr>
<tr>
<td>E14</td>
<td>4</td>
<td>234</td>
<td>65</td>
<td>12</td>
<td>12</td>
<td>8</td>
</tr>
<tr>
<td>Total</td>
<td>NA</td>
<td>2568</td>
<td>790</td>
<td>259</td>
<td>210</td>
<td>67</td>
</tr>
</tbody>
</table>

Experimental Group

<table>
<thead>
<tr>
<th>ID</th>
<th>Stage of Flight</th>
<th>Flight instrument</th>
<th>FTD</th>
<th>Frasca 141</th>
<th>pretest</th>
<th>posttest</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>number</td>
<td>hours</td>
<td>hours</td>
<td>hours</td>
<td>hours</td>
<td>error</td>
</tr>
<tr>
<td>E1</td>
<td>1</td>
<td>82</td>
<td>24</td>
<td>9</td>
<td>9</td>
<td>6</td>
</tr>
<tr>
<td>E2</td>
<td>1</td>
<td>100</td>
<td>26</td>
<td>6</td>
<td>6</td>
<td>2</td>
</tr>
<tr>
<td>E3</td>
<td>3</td>
<td>158</td>
<td>80</td>
<td>31</td>
<td>17</td>
<td>4</td>
</tr>
<tr>
<td>E4</td>
<td>7</td>
<td>259</td>
<td>47</td>
<td>31</td>
<td>17</td>
<td>1</td>
</tr>
<tr>
<td>E5</td>
<td>2</td>
<td>170</td>
<td>100</td>
<td>42</td>
<td>42</td>
<td>2</td>
</tr>
<tr>
<td>E6</td>
<td>1</td>
<td>150</td>
<td>50</td>
<td>20</td>
<td>20</td>
<td>6</td>
</tr>
<tr>
<td>E7</td>
<td>7</td>
<td>320</td>
<td>100</td>
<td>30</td>
<td>16</td>
<td>3</td>
</tr>
<tr>
<td>E8</td>
<td>6</td>
<td>260</td>
<td>40</td>
<td>15</td>
<td>11</td>
<td>7</td>
</tr>
<tr>
<td>E9</td>
<td>3</td>
<td>190</td>
<td>60</td>
<td>13</td>
<td>13</td>
<td>9</td>
</tr>
<tr>
<td>E10</td>
<td>3</td>
<td>188</td>
<td>55</td>
<td>18</td>
<td>18</td>
<td>2</td>
</tr>
<tr>
<td>E11</td>
<td>1</td>
<td>109</td>
<td>29</td>
<td>10</td>
<td>7</td>
<td>5</td>
</tr>
<tr>
<td>E12</td>
<td>3</td>
<td>146</td>
<td>50</td>
<td>12</td>
<td>12</td>
<td>9</td>
</tr>
<tr>
<td>E13</td>
<td>4</td>
<td>202</td>
<td>64</td>
<td>10</td>
<td>10</td>
<td>3</td>
</tr>
<tr>
<td>E14</td>
<td>4</td>
<td>234</td>
<td>65</td>
<td>12</td>
<td>12</td>
<td>8</td>
</tr>
<tr>
<td>Total</td>
<td>NA</td>
<td>2568</td>
<td>790</td>
<td>259</td>
<td>210</td>
<td>67</td>
</tr>
</tbody>
</table>
Data from Table 4.1 shows that the control group as a whole averaged slightly more flight time and slightly less instrument time than the experimental group. Average Frasca 141 FTD time varied from 13.3 in the control group to 15.0 in the experimental group. However, that statistic was skewed by an outlier (ID number E5 was 4 standard deviations from the mean) and when the experimental group mean was calculated excluding that score they averaged 13.3 hours on the Frasca 141 FTD.

Since the control group had one less pilot than the experimental group, a comparison of total error scores between groups was calculated in percent. Figure 4.1 shows the change in total CTM error scores for each group expressed as a percent of total possible errors. The pretest error scores for the two groups were within one percent. The posttest scores showed a difference of 14%. The control group as a whole showed a 9% increase in total CTM errors, and the experimental group showed a 54% decrease in total errors.

![Change in total CTM error for experimental and control groups](image)

Figure 4.1. Graph showing the change in total CTM error scores for each group expressed as a percent of total possible errors. The control group as a whole showed a 9% increase in their total CTM errors, and the experimental group showed a 54% decrease in total errors.
The distribution of types of pretest and posttest CTM errors for each pilot in the control group is shown in Table 4.2 and for the pilots in the experimental group in Table 4.3. The 20 possible CTM errors included seven A/N, seven A/C, five A/N, and one A/A error. The A/A error was included as an example of a prioritization error between two discrete aviate tasks and is included in total error score calculations but not in the following comparison of types of errors between and within groups.

Table 4.2. Distribution of CTM error scores by type for the control group. Statistical data is shown at the bottom of the table.

<table>
<thead>
<tr>
<th>Type</th>
<th>A/N</th>
<th>A/C</th>
<th>N/C</th>
<th>A/A</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total possible</td>
<td>7</td>
<td>7</td>
<td>5</td>
<td>1</td>
<td>20</td>
</tr>
<tr>
<td>C1</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>C2</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>0</td>
<td>7</td>
</tr>
<tr>
<td>C3</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>C4</td>
<td>2</td>
<td>3</td>
<td>1</td>
<td>0</td>
<td>6</td>
</tr>
<tr>
<td>C5</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>C6</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>C7</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>C8</td>
<td>2</td>
<td>4</td>
<td>1</td>
<td>0</td>
<td>7</td>
</tr>
<tr>
<td>C9</td>
<td>1</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>C10</td>
<td>3</td>
<td>4</td>
<td>3</td>
<td>1</td>
<td>11</td>
</tr>
<tr>
<td>C11</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>C12</td>
<td>1</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>C13</td>
<td>4</td>
<td>4</td>
<td>1</td>
<td>0</td>
<td>9</td>
</tr>
<tr>
<td>Total</td>
<td>19</td>
<td>29</td>
<td>11</td>
<td>1</td>
<td>60</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Type</th>
<th>A/N</th>
<th>A/C</th>
<th>N/C</th>
<th>A/A</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total possible</td>
<td>7</td>
<td>7</td>
<td>5</td>
<td>1</td>
<td>20</td>
</tr>
<tr>
<td>C1</td>
<td>0</td>
<td>3</td>
<td>2</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>C2</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>C3</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>C4</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>C5</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>6</td>
</tr>
<tr>
<td>C6</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>C7</td>
<td>3</td>
<td>3</td>
<td>1</td>
<td>0</td>
<td>7</td>
</tr>
<tr>
<td>C8</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>6</td>
</tr>
<tr>
<td>C9</td>
<td>1</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>C10</td>
<td>1</td>
<td>3</td>
<td>3</td>
<td>0</td>
<td>7</td>
</tr>
<tr>
<td>C11</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>C12</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>C13</td>
<td>4</td>
<td>5</td>
<td>1</td>
<td>0</td>
<td>10</td>
</tr>
<tr>
<td>Total</td>
<td>19</td>
<td>32</td>
<td>13</td>
<td>2</td>
<td>66</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Statistic</th>
<th>Pretest</th>
<th></th>
<th>Posttest</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>1.5</td>
<td>2.2</td>
<td>0.8</td>
<td>0.1</td>
</tr>
<tr>
<td>Median</td>
<td>1.0</td>
<td>3.0</td>
<td>1.0</td>
<td>0.0</td>
</tr>
<tr>
<td>Mode</td>
<td>2.0</td>
<td>3.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>Std dev</td>
<td>1.2</td>
<td>1.5</td>
<td>1.0</td>
<td>0.3</td>
</tr>
</tbody>
</table>
Table 4.3. Distribution of CTM error scores by type for the experimental group. Statistical data is shown at the bottom of the table.

<table>
<thead>
<tr>
<th>Type</th>
<th>A/N</th>
<th>A/C</th>
<th>N/C</th>
<th>A/A</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total possible</td>
<td>7</td>
<td>7</td>
<td>5</td>
<td>1</td>
<td>20</td>
</tr>
<tr>
<td>E1</td>
<td>1</td>
<td>3</td>
<td>2</td>
<td>0</td>
<td>6</td>
</tr>
<tr>
<td>E2</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>E3</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>E4</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>E5</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>E6</td>
<td>1</td>
<td>3</td>
<td>2</td>
<td>0</td>
<td>6</td>
</tr>
<tr>
<td>E7</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>E8</td>
<td>0</td>
<td>3</td>
<td>3</td>
<td>0</td>
<td>7</td>
</tr>
<tr>
<td>E9</td>
<td>3</td>
<td>4</td>
<td>2</td>
<td>0</td>
<td>9</td>
</tr>
<tr>
<td>E10</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>E11</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>E12</td>
<td>4</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>9</td>
</tr>
<tr>
<td>E13</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>E14</td>
<td>4</td>
<td>5</td>
<td>1</td>
<td>1</td>
<td>8</td>
</tr>
<tr>
<td>Total</td>
<td>18</td>
<td>31</td>
<td>16</td>
<td>4</td>
<td>67</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Type</th>
<th>A/N</th>
<th>A/C</th>
<th>N/C</th>
<th>A/A</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total possible</td>
<td>7</td>
<td>7</td>
<td>5</td>
<td>1</td>
<td>20</td>
</tr>
<tr>
<td>E1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>E2</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>E3</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>E4</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>E5</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>E6</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>E7</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>E8</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>E9</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>E10</td>
<td>3</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>E11</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>E12</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>E13</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>E14</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>Total</td>
<td>10</td>
<td>14</td>
<td>6</td>
<td>0</td>
<td>30</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Type</th>
<th>A/N</th>
<th>A/C</th>
<th>N/C</th>
<th>A/A</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total possible</td>
<td>7</td>
<td>7</td>
<td>5</td>
<td>1</td>
<td>20</td>
</tr>
<tr>
<td>Median</td>
<td>1.0</td>
<td>2.5</td>
<td>1.0</td>
<td>0</td>
<td>4.5</td>
</tr>
<tr>
<td>Mode</td>
<td>1.0</td>
<td>3.0</td>
<td>1.0</td>
<td>0.0</td>
<td>2.0</td>
</tr>
<tr>
<td>Std dev</td>
<td>1.4</td>
<td>1.4</td>
<td>0.9</td>
<td>0.5</td>
<td>2.8</td>
</tr>
</tbody>
</table>

Mean | 1.3 | 2.2 | 1.1 | 0.3 | 4.8 |
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Median</td>
<td>1.0</td>
<td>1.0</td>
<td>0.0</td>
<td>0</td>
<td>2.0</td>
</tr>
<tr>
<td>Mode</td>
<td>1.0</td>
<td>1.0</td>
<td>0.0</td>
<td>0</td>
<td>2.0</td>
</tr>
<tr>
<td>Std dev</td>
<td>0.8</td>
<td>0.9</td>
<td>0.6</td>
<td>0</td>
<td>1.6</td>
</tr>
</tbody>
</table>
Because there was only one A/A error challenge point in the LOFT, and it represented only 1.0% of total errors in the pretest and 0.7% of total errors in the posttest, that particular error data point was not included in the graphs of error score breakdowns into type. Comparison of pretest error scores (Figure 4.2) shows the experimental group had 5% fewer A/N errors, 10% more A/C errors, and 30% more N/C errors than the control group. Comparison of posttest error scores (Figure 4.3) shows the experimental group had 47% fewer A/N errors, 56% fewer A/C errors, and 53% fewer N/C errors than the control group.

Figure 4.2. Graph showing pretest relationship of type of CTM error between control and experimental group. The experimental group had 5% fewer A/N errors, 10% more A/C errors, and 30% more N/C errors than the control group.
Posttest variation in type of CTM errors for control and experimental groups

<table>
<thead>
<tr>
<th>Type of CTM error</th>
<th>Control group posttest error scores</th>
<th>Experimental group posttest error scores</th>
</tr>
</thead>
<tbody>
<tr>
<td>A/N</td>
<td>35</td>
<td>19</td>
</tr>
<tr>
<td>A/C</td>
<td>30</td>
<td>14</td>
</tr>
<tr>
<td>N/C</td>
<td>25</td>
<td>12</td>
</tr>
</tbody>
</table>

Figure 4.3. Graph showing posttest relationship of type of CTM error between control and experimental group. The experimental group had 47% fewer A/N errors, 56% fewer A/C errors, and 53% fewer N/C errors than the control group.

The distribution or errors for the control group shows there was not a significant change in any one type of error from pretest to posttest (Figure 4.4). The group made the same number of A/N errors in both the pretest and posttest flights, A/C errors increased 9% and N/C errors increased 15%. There was not a direct relationship between the types of errors made by a given pilot in both pretest and posttest. There were a few exceptions, for example pilots C9, C12, and C13 made almost the same number of each type of error in the pretest as in the posttest. As a whole the experimental group showed a 44% decrease in A/N errors, a 55% decrease in A/C errors, and a 63% decrease in N/C errors (Figure 4.5). The A/N error at challenge point 8 was the most common, and the second most common were the A/C error at challenge point 2 and the N/C error at challenge points 4 and 13.
Figure 4.4. Graph showing relationship of each type of error for the control group pretest and posttest flights.

Figure 4.5. Graph showing relationship of each type of error for the experimental group pretest and posttest flights.
4.2 Null Hypothesis Test

As discussed in Section 3.10.1 the Mann-Whitney test was determined to be the most appropriate statistic to use for testing the null hypothesis because the distribution, although normal with acceptable homoscedasticity, had large values at both tails (Figure 3.8). The following null and alternative hypotheses were originally proposed:

\[ H_0: \text{After completing a CTM specific training course, pilots will show no change in task prioritization errors.} \]

\[ H_a: \text{After completing a CTM specific training course, pilots will show a change in task prioritization errors.} \]

If \( D_1 \) and \( D_2 \) represent the sample distributions for the control group and experimental group, respectively, then the test would be stated as:

\[ H_0: D_1 = D_2 \]
\[ H_a: D_1 \neq D_2 \]

The Mann-Whitney was calculated as described by McClave and Sincich (2003). The raw data from the two samples were first combined into a set of \( N = n_1 + n_2 \) elements, which were then ranked from lowest to highest, including tied rank values for six of the scores (Table 4.4). The rankings were then re-sorted into the two separate samples for analysis.
Table 4.4. Table showing raw scores and rankings used in the Mann-Whitney calculation. The raw scores are the pretest-posttest distribution of difference in error scores for the control group (D₁) and the Experimental group (D₂).

<table>
<thead>
<tr>
<th>Count</th>
<th>Raw Scores (D₁)</th>
<th>Raw Scores (D₂)</th>
<th>Ranks (D₁)</th>
<th>Ranks (D₂)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3</td>
<td>-4</td>
<td>25</td>
<td>5.5</td>
</tr>
<tr>
<td>2</td>
<td>-3</td>
<td>0</td>
<td>8</td>
<td>19</td>
</tr>
<tr>
<td>3</td>
<td>-1</td>
<td>1</td>
<td>14.5</td>
<td>22</td>
</tr>
<tr>
<td>4</td>
<td>-1</td>
<td>-2</td>
<td>14.5</td>
<td>10.5</td>
</tr>
<tr>
<td>5</td>
<td>5</td>
<td>0</td>
<td>26</td>
<td>19</td>
</tr>
<tr>
<td>6</td>
<td>-1</td>
<td>-3</td>
<td>14.5</td>
<td>8</td>
</tr>
<tr>
<td>7</td>
<td>6</td>
<td>-3</td>
<td>27</td>
<td>8</td>
</tr>
<tr>
<td>8</td>
<td>-1</td>
<td>-7</td>
<td>14.5</td>
<td>3</td>
</tr>
<tr>
<td>9</td>
<td>0</td>
<td>-9</td>
<td>19</td>
<td>1</td>
</tr>
<tr>
<td>10</td>
<td>-4</td>
<td>2</td>
<td>5.5</td>
<td>24</td>
</tr>
<tr>
<td>11</td>
<td>1</td>
<td>-1</td>
<td>22</td>
<td>14.5</td>
</tr>
<tr>
<td>12</td>
<td>-1</td>
<td>-8</td>
<td>14.5</td>
<td>2</td>
</tr>
<tr>
<td>13</td>
<td>1</td>
<td>-2</td>
<td>22</td>
<td>10.5</td>
</tr>
<tr>
<td>14</td>
<td>-5</td>
<td></td>
<td></td>
<td>4</td>
</tr>
</tbody>
</table>

Six of the ranked scores resulted in ties (t = 14.5). Siegel (1956) noted that in the Mann-Whitney test ties have only slight effects, even when a large proportion of the scores are tied. There is a correction for ties that can be applied to the standard deviation of the sampling distribution. However, Seigel (1956) stated that correcting for ties will increase the value of the z score, making it more significant. Thus, if ties are not corrected then the statistic will be slightly more conservative in that the value of p will be slightly larger. For this calculation the correction for ties was not used.

When using the Mann-Whitney test, if the two sets being ranked have unequal numbers of data, then the test statistic must be ranked on the set with the smaller value.
of \( n \) (McClave & Sincich, 2003). Because \( n_1 < n_2 \) (\( n_1 \) for the control group = 13 and \( n_2 \) for the experimental group = 14) the statistic was based on \( n_1 \). The Mann-Whitney U value was calculated using SPSS 11.0 software (Table 4.5) which indicated the null hypothesis could be rejected in favor of the alternative hypothesis at a probability level \( p = 0.029 \).

Table 4.5 Results from SPSS calculation of Mann-Whitney test.

<table>
<thead>
<tr>
<th>Mann-Whitney U test - Not corrected for ties</th>
</tr>
</thead>
<tbody>
<tr>
<td>pretest-posttest difference</td>
</tr>
<tr>
<td>U</td>
</tr>
<tr>
<td>U critical</td>
</tr>
<tr>
<td>Z</td>
</tr>
<tr>
<td>p (2-tailed Sig)</td>
</tr>
</tbody>
</table>
5. DISCUSSION AND CONCLUSIONS

The main objective of this study was to design a CTM training course consistent with mainstream flight training practices and determine if that training affected pilots' task prioritization performance in simulated flight. A secondary objective was to observe and comment on pilot behaviors related to multitasking processes, specifically CTM theory. Third, a critique of the study's strengths and limitations helped suggest criteria for further investigation.

5.1 Training Effect

Data presented in Chapter 4 show the control group as a whole made the same or more prioritization errors overall in the posttest flight compared to the pretest. Individual pilots showed an increase, a decrease, or no change in errors (Tables 4.1 and 4.2). Such a distribution would be expected from randomly sampling a group of pilots during two discrete flights. If there were no effect from the CTM training course then the experimental group should show a similar distribution of pretest and posttest scores. However, the experimental group had a large decrease in total CTM errors between pretest and posttest flights compared to the control group (Tables 4.1 and 4.3 and Figure 4.1). If measures of control for internal validity were met, it follows that the CTM training course did cause a reduction in task prioritization errors for pilots in the experimental group during this study.

It seems reasonable that any well designed training course would show some effect during the short term. However, a major question that arises is, how long will it take that effect to disappear, or to drop below acceptable performance standards? The answers to those questions would need to be assessed by testing the same participants at a later date, as well as controlling for effects of extraneous variables that might affect their performance. As discussed in section 3.6.8 the amount of time between the pretest and posttest simulated flights (two weeks) represented a trade-off between
internal and external validity. The time period was kept short enough to reduce history effects, but that did not allow the study to comment on longer term effects of the training. Because pretest data described in section 3.3 indicated no correlation between this group of participants' total flight time, instrument time, or FTD time and their CTM performance, those particular extraneous variables might not have a large influence on future CTM performance for those pilots. However, control for other extraneous variables over a longer period of time, which would include further training in human factors and additional flight experience, might be difficult.

Another viable question is whether or not any learning actually took place. This investigator can only comment that pilots who received training showed a decrease in CTM errors and an improvement in performance over a two week period of time, but cannot necessarily make an assumption that the pilots actually retained the new information or learned new behaviors that will endure. One of the most widely used theories of learning and retention is the dual memory model as described by Schunck (2004) and adopted by the FAA in their flight instructor training literature (FAA, 1999). According to the dual memory model, information is processed through inputs (primarily visual and auditory) to the sensory register. In order to transfer information to the long term memory the learner must relate incoming information to concepts and ideas already in memory. Since all the pilots in the experiment had previously studied concepts of prioritization and task management during their regular flight training it is possible that the reduction in CTM errors by the experimental group might represent a sensitization effect. The only difference between the two groups might have been that the experimental group was focused on those concepts during the short term and did not actually code the information into their long term memory. The issue of whether learning occurred is a critical one and also difficult to resolve because a teacher or instructor often does not have the ability to evaluate students after they leave the learning environment. More follow-up studies are needed to comment on the long term effects of the training. Additionally, a qualitative response from participants at some future time might also reflect on whether or not they believed learning occurred.
Pilots in the experimental group who showed the greatest reduction in CTM error scores were the ones that originally made the most errors (Table 4.1). It could be that the reduction in errors might simply represent a regression toward the mean for those pilots. However, the fact that several pilots in the control group also scored a large number of errors in the pretest without a corresponding reduction in errors for the posttest indicates that regression was probably not the cause for that trend in the experimental group. What the data suggests is that pilots who performed the worst seemed to benefit more from the training than those who initially made a low number of errors. Alternatively, pilots who made only one or two errors in the pretest and posttest were not able to be evaluated with respect to a training effect since there were only a fixed number of challenge points and it was not possible to show a large improvement in error scores for those pilots.

As part of the CTM training course pilots in the experimental group wrote a reflection of their perceived task prioritization performance during their most recent flight and some debriefed that flight verbally with the investigator. Many pilots noted that they had a heightened awareness of the highest priority item, flying the airplane, during those flights but they did not change the way they conducted flow checks or checklists. Rather, the training reaffirmed their knowledge and heightened their awareness of aviate tasks as being the most important. One pilot wrote that the entire instrument flight went more smoothly because he kept reminding himself to fly the plane first. Another student wrote the following in her reflection:

I did try and apply what we talked about and found myself falling into my old habits on the first flight I took and then on future flights I made myself slow down, take a deep breath, and fly the plane first and answer the phone later. I felt that the flight went much more smoothly once I got my priorities straight. I was far less stressed too. I will make a point to aviate, navigate, communicate and in that order on future flights.

(with permission, E. Jakubowicz, written communication, February 5, 2005)

The hope is that because pilots in the experimental group were given the opportunity to reflect and apply concepts from the CTM short course that they were able to process and retain the information and that learning occurred.
5.2 Multitasking Behaviors

As discussed in Section 2.2., different cognitive models have been proposed in past and recent literature to explain how pilots process more than one concurrent task. The challenge points in the LOFT (Appendix D) were designed to simulate situations that might occur in a routine flight where pilots were confronted with two or more simultaneous tasks that had to be prioritized.

5.2.1 Single Channel Theory

To resolve some of the LOFT challenges, some pilots made prioritization errors that could be described by single channel theory (SCT) which posits that a person will abandon one task until the other task is completed (Broadbent, 1958, Welford, 1967). For example, more than half the pilots in both groups made errors at challenge point 8 (Appendix D) which involved executing the MAP for the ILS 13R at Boeing Field. The procedure calls for the pilot to climb via the IBFI localizer course to 2000 feet, then to identify COGAR intersection to commence climb to 5000 feet while continuing to track the localizer course (Appendix C).

At challenge point 8 some pilots completely ignored the aviate task while they were executing the navigate task. Many of the pilots became fixated on the task of either programming the GPS for the waypoint or tuning and identifying the VOR to identify the cross radial for COGAR and either strayed off course, deviated from altitude, or both, while attempting to identify the fix. In several cases the video tapes showed pilots were not even looking at their flight instruments but had leaned over to the right side of the cockpit to view and input the GPS unit more directly. A few pilots who did that were off altitude by as much as 500 feet, off the localizer course by a full needle deflection, or off heading by as much as 90 degrees within a matter of only a few seconds. Additionally, some of those pilots did not immediately recognize their altitude and course/heading deviations when they looked back at the flight.
108

instruments. Moray and Rotenberg (1986) called that phenomenon "cognitive lockup" which in the situation at challenge point 8 meant the pilots became attentionally locked into the navigate task and did not notice other large magnitude deviations. Moray and Rotenberg (1986) concluded the cognitive lockup behavior as evidence that people deal with problems serially rather than switching between tasks.

The FAA refers to that type of attentional tunneling as fixation (FAA, 2001; Jeppesen, 2003a). A fixation error means the pilot puts all his attention on one task to the exclusion of others, which is consistent with SCT. Although the pilots in the study were familiar with the GPS unit from previous use in both the airplane and FTD, they still became fixated when programming the waypoint in flight. Pilots who pre-programmed the GPS while still on the ground at Renton did not make the fixation error at challenge point 8.

The issue of fixation has become an area of great concern in the flight training industry in recent years. Over the past five years general aviation cockpits have incorporated more sophisticated IFR certified GPS units, and in the past three years the new flat panel digital primary flight displays (PFDs) and multifunction displays (MFDs) have been installed in training aircraft. Wilson (1998) found that as the level of sophisticated instruments and automation increases on airline flight decks the potential for CTM errors also increases. It is likely that the same potential exists for increased sophistication in general aviation cockpits, including training aircraft.

5.2.2 Single Resource Theory

Some of the pilots showed the ability to share resources between two tasks simultaneously, which is the basis of single resource theory (SRT; Kahneman, 1973; Lindsay & Norman, 1972) model of cognitive processing. Wickens (2002) posited that mental resources can be allocated to different tasks and if one is simple (or automatic) it requires almost no mental resources, so more of the available resources can be allocated to the other task. An example of that was observed at challenge point
7 (Appendix D) which required the pilot to execute a missed approach procedure (MAP) or go-around. During a MAP the most critically important aviate task is to initiate a positive rate of climb before doing any other task. Participants in this study had been previously trained to use a mnemonic memory aid for the MAP called “five Cs” which is designed to guide them in doing tasks in the order of higher to lower priority. The five Cs stands for “cram, climb, clean, cool, call” and means the pilot should cram the throttle (put the throttle, and propeller if applicable, to the full forward position), initiate a positive rate of climb, clean up the airplane aerodynamically (retract the landing gear and flaps), cool the engine via opening of cowl flaps or richening of fuel mixture, and call ATC as last priority.

The first two actions (pushing the throttle forward while pitching for best rate of climb attitude) were done simultaneously by most of the pilots at challenge point 7. According to SRT, the time for the pilot to execute both those tasks simultaneously will be less than the time required to do them sequentially because the pilot is able to perform the more automatic task (adjusting the throttle) without using much mental effort so they can concentrate more on the more difficult task of pitching to the correct attitude. According to Wickens (2002) there is some task interference, but it results not from postponement of one task over another (as predicted by SCT) but rather from the concurrence of the tasks. Wickens stated that both tasks will probably be executed in the amount of time of the longest task; e.g. if it takes three seconds to push the throttle forward and eight seconds to pitch for climb attitude, the total time for both tasks will be eight seconds (the time it takes for the more difficult task). SCT would predict a total time of 11 seconds for the pilot to execute the tasks sequentially. Because most pilots did both tasks simultaneously in a relatively short time they were most likely sharing resources between the tasks according to SRT.

An exception was pilot C9, who did not use the memory aid and made the same prioritization error at challenge point 7 during both the pretest and posttest flights. Instead of putting the throttle to full power and initiating a climb, pilot C9 put the throttle to about half power, stopped and called ATC, then put the throttle to about the three-quarter position. Then he stopped and switched radio frequencies (during
that time in the pretest flight the climb degraded and the plane went into a descent, in the posttest it leveled off), after which he called ATC, then finally put the throttle to full forward and established a climb. The pilot was having great difficulty switching between tasks and did not seem to be able to timeshare between tasks. In fact the whole sequence took over two minutes in the pretest and almost the same amount of time in the posttest before the pilot actually established a positive rate of climb. Other pilots typically achieved a positive rate of climb within 10-15 seconds.

The MAP scenario just described is probably one of the most critical to flight safety, since on a typical MAP from an ILS the aircraft is only 200 feet above the ground, and if a climb is not initiated immediately there is a high risk for a controlled flight into terrain (CFIT) type of accident. Because pilot C9 made the same error in a similar manner both times, it could represent a learned behavior that would require significant additional training effort to rectify, or possible it could represent an inability to process the tasks cognitively, which could be potentially dangerous.

5.2.3 Multiple Resource Theory

As discussed in section 2.2.3, parallel or concurrent processing of two or more tasks (time sharing) is an important concept of multiple resource theory (MRT) (North, 1977; Wickens, 1980) in that it assumes parallel or concurrent processing is possible. Additionally, tasks that do not compete for the same resources, such as a visual task and an auditory task, are easier to perform simultaneously than two tasks that use the same resources (e.g.: two visual/spatial tasks) (Liu & Wickens, 1992; Wickens et al., 2003). Observations from this study confirmed that the visual and auditory tasks interfered less with one another than did two visual tasks. That was demonstrated by comparing the pilots’ behavior and task prioritization performance at challenge points 2, 4, and 5 (Appendix D).

At challenge point 2 the pilot had to prioritize between a higher priority aviate task and a communicate task. The aviate task was visual. At point 2 the pilot had to
visually interpret input from the altimeter, attitude indicator, and vertical speed indicator to level the airplane at a designated altitude. For the communicate task the pilot had to attend to an incoming radio call from ATC (auditory task). Challenge point 4 was similar, except that the visual input was a navigate task (interpreting movement on the VHF navigation display and intercepting the localizer course) and the communicate task again involved attending to an incoming radio call from ATC (auditory task). Most pilots were able to level off the airplane at challenge point 2 or make the turn to intercept at challenge point 4 and attend to the incoming call at the same time, thus effectively time sharing their attention between tasks that required different cognitive resources as suggested by MRT. In the pretest seven of the 27 pilots made the A/C prioritization error at challenge point 2 and three pilots made the N/C error at challenge point 4. In the posttest three pilots made the A/C error and three made the N/C error. A few pilots postponed the communicate task until after completing the aviate or navigate task, which could indicate they were processing inputs in a more sequential manner that would be consistent with SCT.

In contrast to challenge points 2 and 4, challenge point 5 involved a conflict between a visual aviate task (flying the aircraft at a constant airspeed and descent rate) and a visual navigate task (tracking the localizer and determining that the glideslope had failed). According to MRT two tasks that use the same resources are more difficult to perform simultaneously. Indeed, 11 of the 27 pilots made the prioritization error at challenge point 5 in the pretest, and six pilots made the error in the posttest. Interestingly, pilots in the experimental group showed greater posttest improvement at challenge point 5 than at the other A/N challenge points.

MRT could also explain the situation at challenge point 8 discussed in section 5.2.1. The videotapes clearly showed that many pilots completely ignored the aviate task to lean over and focus on the GPS, as described earlier. However, some of the pilots could have been unsuccessfully attempting to time share or switch between tasks, which would be better described by MRT rather than SCT.
5.2.4 Workload and Attention Management

The LOFT was carefully designed to be as realistic as possible, and many of the CTM challenge points required little or no intervention by the experimenter, rather the pilot had to make prioritization decisions during normal flight operations. The LOFT placed the pilot in a relatively high workload environment at different times throughout the flight. During any flight the highest workloads are during takeoff and departure and again during approach, and landing (FAA, 1991). The FAA suggests that pilots should conduct as many tasks as possible during low workload times, and Raby and Wickens’ (1994) empirical studies showed that pilots who conducted tasks during low workload periods performed better. For example, pilots can set up navigation radios, check weather, and brief approach procedures before entering the approach environment. Although pilots are taught to manage time and perform duties during low workload times, the investigator has observed many times that during low workload periods pilots become complacent and do not stay ahead of the airplane with regards to obtaining current weather reports and setting up navigational radios, etc. It is possible they become bored from lack of stimulation during those times.

During the pretest flights in this study some pilots accomplished tasks well ahead of time, and others waited until their workload had already increased to perform routine tasks. Although conducting tasks at low workload times was not directly related to any specific challenge point, pilots who completed as many tasks as possible ahead of time had better CTM performance during the high workload times (departure, approach, and missed approach). For example, pilots who pre-programmed the GPS while still on the ground at Renton did not make the fixation error at challenge point 8 as discussed in section 5.2.1. Likewise, pilots who obtained weather updates well prior to commencing the approach, and who briefed the approach or hold well ahead, had more time for the actual conduct of those maneuvers in addition to being able to monitor flight and engine instruments.
The CTM training course emphasized using low workload times to conduct non-critical tasks ahead of time and reminded pilots to constantly be checking the status of the aircraft and preparing for the next phase of the flight. The concept was mainly stressed through procedural discipline and attendance to checklists and flow checks. Proper use of procedures, flow checks, and checklists increased the pilots’ ability to manage high workload times by reducing the workload, and it also helped pilots to perform tasks concurrently, as discussed in the next section.

5.2.5 Concurrent Task Management

Because a substantial amount of research exists regarding task management and concurrent task management (Funk, 1991; Funk et al., 2003; Raby and Wickens, 1994; Rogers, 1996; Schutte & Trujillo, 1996; Wickens et al., 2003) this study used criteria from previous studies to design elements of the CTM training course and to note how particular CTM error performance changed between pretest and posttest flights. Funk et al. (2003) showed that the two factors pilots identified as influencing task prioritization were the status of the task and procedure (Figure 2.1). Therefore, the CTM training course strongly emphasized procedural discipline, including strict conduct of checklists, briefings, flow checks, and mnemonic memory aids at appropriate times. The training course did not stress any one type of CTM error over another but rather addressed task prioritization concepts with respect to various scenarios and possible types of errors related to improper use of checklists or lack of procedural discipline, as well as focusing on the ANC scheme to assess what took priority in a given situation.

During the posttest pilots in the control group did not show much change in the way they conducted flow checks and standard procedures, but several pilots in the experimental group showed much improved procedural discipline in the posttest flights. Specifically, pilots E2, E5, E7, E8, E9, E11, E12, and E14 each used their checklists more vigilantly and performed briefings and flow checks well ahead of time
low workload periods in the posttest compared to the pretest flights. Pilots E2 and E5 had only a 10% CTM error rate in the pretest and did not show a change in errors in the posttest. However, Pilots E7, E8, E12, and E14 showed a change from as high as 45% error to zero error between pretest and posttest flights (Table 4.3). Therefore, it could be that checklist vigilance and procedural discipline were large contributors to reduction in CTM errors.

An good example of how pilots proper use of flow checks and procedure to handle concurrent tasks was at challenge point 12 (Appendix D). That challenge point involved an A/A task in which the instructor caused carburetor icing while the pilot was in cruise flight, which caused a steady loss of power. Most of the pilots noticed the reduction in RPM and conducted the proper flow check to troubleshoot the fuel, ignition, and induction systems, which included addition of carburetor heat (thus solving the problem). During that troubleshooting process, pilots who followed the flow check did not deviate from altitude, airspeed, or course while troubleshooting the problem. Without exception, the five pilots in the pretest and two in the posttest who made the A/A error at challenge point 12 did so because they did not follow their flow check procedure. Those pilots they either got slow or descended because they either did not determine the cause of the power loss or they did nothing. One pilot in the pretest (E12) experienced a total loss of power and declared an emergency (after that point the instructor removed the carburetor icing and the pilot restarted the engine and resumed the flight so that the remaining challenge points could be addressed). Future experiments might include more such tasks to comment further on the relationship of procedural discipline and CTM performance.
5.3 Conclusions

Experimental analysis showed that the group of university flight students who participated in the CTM short course improved their task prioritization performance over a two week period of time. Those pilots showed improvement in each of the three main types of prioritization errors. It was not determined whether that performance increase had a longer lasting effect. Pilots who did not participate in the CTM short course did not markedly improve their prioritization performance. Those pilots showed either an increase, decrease, or no change. Based on the control group's posttest performance there did not seem to be a practice effect from the pretest.

Pilots in the experimental group who made the most CTM errors in the pretest showed the most improvement after participating in the CTM training course. However, that decrease in CTM errors did not seem to be a result of regression toward the mean since control group pilots with comparable pretest performance showed negligible change in posttest flights. Experimental group pilots who made only one or two CTM errors in the pretest might have improved or not, but that criteria was undeterminable since there were a fixed number of challenge points from which data were collected instead of all assessing all possible errors made.

Pilot behavior was described by different cognitive processing models with regards to various behaviors and CTM errors. Some behaviors could be described by single channel theory (SCT) wherein the pilot did not seem to be able to process inputs and perform tasks simultaneously. Single resource theory (SRT) described other behaviors in which pilots seemed to be able to time share or switch between tasks. As described by multiple resource theory (MRT), tasks that shared the same cognitive resources, such as two visual tasks, were not performed as well together as tasks that did not, such as a visual task and an auditory one. Additionally, auditory inputs such as a call from ATC were more interrupting than visual tasks, such as tracking a localizer course or a change in engine instrument readings. One exception was that of pilots fixating on the GPS display to the exclusion of aircraft control, sometimes showing dangerously large deviations in altitude and course. Fixation errors are of
critical importance in the current flight training environment as modern cockpits utilize more sophisticated displays and avionics.

Pilots who practiced a high level of procedural discipline and proper use of checklists, flow checks, and mnemonic memory devices performed better than those who did not, especially when they were able to perform routine tasks during low workload periods. Experimental group pilots who improved their procedural discipline and performed more tasks during low workload periods showed a marked improvement in CTM performance in posttest flights.

5.4 Limitations

Although many of the recommendations from previous studies were addressed by this experimental analysis, other items arose during the course of the study as limiting factors. Originally the hope was to use a more sophisticated experimental design, such as a Solomon four group, but there were not enough participants. In fact, mortality issues (section 3.6.5) included participants withdrawing due to illness or personal reasons as well as corrupted data from a deviation in the experimental procedures. The latter should be addressed in future studies through more thorough supervision and training of instructors.

As discussed in section 3.6.8, the time period between pre and posttest FTD flights was a compromise between internal and external validity. The two week duration was short enough to increase internal validity because it minimized possible history effects from extraneous learning, but it could also have made the results less meaningful with regards to the long term effects of training. Future studies might examine whether such training has longer term effects.

Selection of the participants was another compromise between internal and external validity. Participants comprised a relatively homogeneous group with respect to previous training and experience, but those same criteria means that results may not be generalized to a more variable group of general aviation pilots.
The study collected data for a fixed number of errors (20 errors at 14 challenge points). However, many pilots made prioritization errors at times during the flight other than at the challenge points. The technique used in the pilot study of gathering error data throughout the flight was much more difficult and necessitated a highly experienced instructor with an in depth understanding of CTM theory, which made it infeasible for this study. However, it is possible that by limiting observations to a set number of challenge points the study was creating a false ceiling effect such that pilots could show a zero error score when in fact they had made errors. It could be a difficult challenge to design an experiment that analyzed all possible errors made throughout the flight.

The FTDs used in the study were fully FAA approved and equipped with standard analog flight and engine instruments as well as standard avionics. The Garmin GNS430 approach certified GPS was a more advanced system. Because fixation on the GPS was a major cause of CTM error for more than half the pilots, any experiment such as this one using basic analog instrumentation might not address the challenges faced by pilots using the newest flat panel digital PFD and MFD displays now being installed in many new aircraft.

The training course was carefully designed and delivered to be consistent with current practices used by FAA certified flight instructors and flight schools. However, the investigator has a substantial amount of experience teaching in the classroom and in airplanes and simulators, as well as experience training flight instructors. Therefore, the experiment could have been only testing one specific instructor and thus not be able to generalize results to other instructors or learning environments.
5.5 Recommendations

Based on the limitations just discussed, several recommendations for future research are suggested. First, the same experiments could be conducted with students at a different flight school. External validity issues could be addressed by comparing results between pilots who have different training backgrounds. If results were similar, then the training course could be more generalized to other training programs. Additionally, the training course should be conducted by other flight instructors with varying levels of experience so that the results can be correlated as to whether or not they were dependent on a particular instructor conducting the training. A less homogeneous group of pilots might also be used to investigate whether pilots with varying levels of experience showed different training effects and if there were some specific level of experience at which pilots showed the greatest effect. External validity could also be enhanced by using a larger sample size and a more powerful experimental design, such as a Solomon four-group. Other experimental designs, such as a time-series design, could be used to determine longer term training effects, either with single subjects or a group of pilots.

The experiment could be conducted with a longer time period between pretest and posttest flights and controlled for extraneous variables to test for long term training effects. Alternatively, these same pilots could be tested at a later date to determine longer term training effects of this study's training course. Qualitative studies could also be used to enhance experiments, such as gathering responses from participants to discern the extent of their learning.

A more sophisticated method to determine all errors during the flight, rather than just at designated challenge points, would be difficult, but could be used to determine if there was a ceiling effect for CTM error rates. To investigate the issues of task fixation as it relates to cockpit complexity and automation, a study could be designed to test pilots in cockpits with various levels of complexity, for example using one of the many new flat panel PFD/MFD or virtual 3D displays installed in many new general aviation aircraft.
REFERENCES


APPENDICES
Human Subjects approval was through the Human Subjects Review Committee (HSRC) at Central Washington University as the principle institution, with Oregon State University approved as the secondary institution. This appendix includes the following documents, which have been altered from their original form only to the extent to meet formatting requirements:

- Approved informed consent form that each participant signed
- Approved recruitment flyer used to advertise the study
- Approved pretest questionnaire used to gain information about each pilot’s experience and flight time.

All clipart is copyrighted by Microsoft Corporation and used here in compliance with the End User License Agreement as stated in Microsoft Word 2002 License number 54185-641-0434003-17881 Section 1 paragraph 5.
CENTRAL WASHINGTON UNIVERSITY

Consent for Research Participation

CONCURRENT TASK MANAGEMENT IN THE COCKPIT: EXPERIMENTAL ANALYSIS OF TASK PRIORITIZATION TRAINING FOR GENERAL AVIATION PILOTS

Principal Investigator: Amy L. Hoover, M.S., Assistant Professor of Flight Technology, Industrial and Engineering Department, College of Education and Professional Studies, Central Washington University (Phone: 509-963-2300).

Co-investigator: Dale R. Wilson, M.S., Associate Professor of Flight Technology, Industrial and Engineering Department, College of Education and Professional Studies, Central Washington University (Phone: 509-963-2298).

24-hour Emergency Contact: 509-899-5178 (Amy Hoover can be reached at this number in case of an emergency).

INVESTIGATOR'S STATEMENT

PURPOSE AND BENEFITS

You are invited to participate in an experimental study to analyze the effectiveness of a task prioritization training program for general aviation pilots. The purpose of this research is to gather data regarding pilot performance in a multitasking environment and to investigate the effectiveness of cockpit task prioritization training. There are possible benefits to you if you decide to participate in this study, which include flight time on the Frasca 141 flight simulator as well as participation in the task prioritization training, which could help you to perform your flight tasks more effectively and reduce errors. There may be some benefits to the flight training community and to society in general. For example, flight instructors and aviation educators may gain knowledge about how to better train pilots to prioritize tasks more effectively in the cockpit, reduce errors, and maximize flight safety. There may be other benefits we do not yet know about.

PROCEDURES

General Information: You have been recruited for participation in this study because you are a Flight Technology major at CWU and are enrolled in at least Instrument Stage II of your flight training. This will be an experimental study involving approximately 40 students. It is important for you to understand the following:

1. If you choose to participate in this study, any personal information you provide to members of our research team, which includes your flight performance
during two simulated instrument flights on the CWU Frasca 141, will not be shared by us with anyone else, including other faculty or your flight instructor(s).

2. If you choose to participate in this study your performance will not affect your grade in any class, including flight labs, in which you are currently enrolled or will be enrolled in the future.

3. If you choose to participate in this study you will not be paid or receive any type of compensation. You will be able to log the two Frasca 141 flights as simulated instrument time in your pilot logbook.

What you will be asked to do: If you choose to participate in this study, you will be asked to do the following things:

1. You will be asked to complete a short questionnaire in which you provide information about your FAA certificates held, your total flight time, total instrument time, and total Frasca 141 time.

2. You will be asked to fly an approximately 1 hour flight in the Frasca 141 using normal instrument procedures for a flight in the Seattle Class B airspace. During this flight the investigator will observe and videotape your performance to analyze task prioritization errors.

3. About two weeks later you will be asked to fly an approximately 1 hour flight in the Frasca 141 using normal instrument procedures for a different flight in the Seattle Class B airspace. During this flight the investigator will observe and videotape your performance to analyze task prioritization errors. During the time between the flights you may be asked to do either a) or b) below:
   a. You will not be asked to do anything additional.
   b. You will be asked to participate in two one-hour learning sessions. The first session will consist of reading selected materials related to cockpit task management and a guided discussion regarding factors of cockpit task management. The second session will consist of a self-study of strategies to improve pilot multitasking performance and a guided discussion of the outcomes.

4. After you fly the second flight and at the end of the study, if you were given the option a) above and did not participate in the training course, you will have the opportunity to take the training course if you desire.

The reason for making a videotape of your simulator flights is to obtain a record of the pilot errors we are analyzing so we can confirm the real-time observations that will be made during your flights. Before we review the videotape for data confirmation, it will be stored in a locked file cabinet. Only the principal investigator, Amy Hoover, will have access to the videotape. Upon completion of the study the videotape will be physically destroyed. All identifying information (e.g., your name) will also be removed from the written data and report and replaced with a numerical identifier that is encoded to identify which study group you are in and what level of instrument experience you have. At the bottom of this form there is a space in which you will be asked to sign your name if you agree to have your flight videotaped, otherwise, we will have to rely only on notes taken during your flight.
It is important for you to understand that your participation will consist of activities identical to your normal flight training activities. These will consist of the two flights in the Frasca 141 simulator, and might include the two learning sessions which are similar in nature to ones you would normally encounter in your everyday ground school courses. Also, it is important for you to understand that these activities constitute only minimal risk, inconvenience, or discomfort, and that the probability and magnitude of harm, inconvenience or discomfort anticipated in this study are not greater than those encountered in daily life. If you permit us to videotape your simulator flight we will ask you to sign a separate section on this consent form that specifically gives permission to make this video recording.

Confidentiality. To protect your privacy, all information we learn from you will be stored in a numerically coded form so that it does not include information that can be used to identify you. Only members of the research team will have access to the research data. If you agree to have your simulator flights videotaped, the tape will be stored in a locked file cabinet until we review it for data confirmation. Only the principal investigator, Amy Hoover, will have access to the videotape. The videotape and all other information collected on you will be physically destroyed at the completion of the study.

Adverse Events. No adverse events are anticipated in this study. In the event that some harm occurs, please speak to the research team immediately. The research project has no funds for treatment or compensation. As a university student, however, you may be eligible for limited health services through student services offices on your campus.

Other Information. You are free to refuse to participate in this study. Your choice about the study does not in any way affect your status as a student at this university, your grade in any course, or your flight progress in flight labs. If you decide to participate in this study, you should know that you have the right to withdraw from the research at any time without penalty to you or loss of benefits to which you are otherwise entitled.

_____________________________  ________________________________
Signature of Investigator        Date
SUBJECT'S STATEMENT

The procedures above have been explained to me and I voluntarily consent to participate in this research activity. I have had an opportunity to ask questions before consenting and future questions that I may have about the research will be answered by one of the investigators listed at the top of this form. If I have questions or concerns about my rights as a research subject I can also call the Central Washington University Human Protections Administrator at (509) 963-3115. I may withdraw from this study at any time without compromising my access to university services I could otherwise normally receive if not involved in this study. In signing this consent form I am not waiving any legal claims, rights, or remedies. A copy of this consent form will be given to me.

Signature of Subject ____________________________ Date ______________

The purposes and procedures for videotaping my Frasca 141 flights have been explained to me and I voluntarily consent to have them videotaped by the research team for these purposes. I have had the opportunity to ask questions before consenting and future questions I might have about the videotaping procedures or my rights as a subject will be answered by one or the investigators listed at the top of this form. I have the right to review the videotape and to request that any or all portions of it be destroyed prior to its being analyzed for this study. If I have questions about my rights I can also call the Central Washington University Human Protections Administrator at (509) 963-3115. I may withdraw from the videotape recording procedure at any time without compromising my access to any courses, flight labs, or other university services I could normally receive if not involved in this study.

Signature of Subject ____________________________ Date ______________

Copies to: Subject

Investigator's file
Volunteers Needed for Research Study

Participants are needed for a study of pilot multitasking

Beginning July 2004 through June 2005

During the study, you will fly two one-hour flights on the Frasca 141 and may participate in an additional training course specific to improving pilot multitasking performance.

You must have completed your Private Pilot License and be enrolled in Instrument Stage II or higher.

For more information contact:

Professor Hoover at 963-2300

Flyer used to recruit participants. The flyer was posted in the CWU Flight Technology building and at the Bowers Field Airport.
Preflight Questionnaire

Central Washington University
Industrial and Engineering Technology Department, Flight Technology

Concurrent Task Management in the Cockpit: Experimental Analysis of Task Prioritization Training for General Aviation Pilots

Conducted by:
Amy L. Hoover, Assistant Professor – principle investigator

Pre-experiment questionnaire:

Date: ____________________  Participant I.D. number ________

Flight Time:

Total flight hours: __________  Total instrument hours: __________

Total hours, Flight Training Device: ________

Total hours, Frasca 141: __________

Stage of flight training currently enrolled:

Instrument:  Stage II _______  Stage III _______

Commercial: Stage IV _______  Stage V _______  Stage VI _______

CFI Stage I _______  Stage II _______  Stage III _______

Multi Stage I _______  Stage II _______
Following is the syllabus for the training course conducted in the pilot study and in the main study. The syllabus has been altered from its original format to meet formatting requirements as well as removal of some figures due to possible copyright infringement issues.

All clipart is copyrighted by Microsoft Corporation and used here in compliance with the End User License Agreement as stated in Microsoft Word 2002 License number 54185-641-0434003-17881 Section 1 paragraph 5.

Specific training course materials are available on request. A description of activities is presented in Section 3.4.
Concurrent Task Management
Training to prioritize in the cockpit

A two week short course
Times and dates to be announced

Instructor: Amy Hoover
Office: RM. 116, Flight Technology Center
Office Hours: As posted or by appointment.
Phone: 963-2300   email: hoovera@cwu.edu

Prerequisites: Instrument Stage II

Course Description and Objectives:
This is a highly interactive course designed to assist you in synthesis of fundamental concepts and practices you can use to build the foundation for your career as a pilot. You will develop your ability to learn and apply knowledge and skills to improve your flying in many ways. You are encouraged to actively participate in your own learning and to construct your own insights by building on the experience you bring with you to this course. You will also benefit by relating what you learn to your own life and to all your flight training. Much of the learning in this course will come from your peers; thus you are encouraged to gain insight from collaboration and cooperation with your classmates and to participate in all class discussions and learning experiences. As your facilitator, I will help you to construct your own learning experiences and to set your personal goals for learning outcomes.
Course Rationale:
You will work to increase your understanding of factors related to workload management, situational awareness and concurrent task management and apply it toward more effectively prioritizing cockpit tasks. Emphasis will be on standard procedures, use of checklists, proper flow checks, and your ability to quickly and effectively assess and prioritize tasks in both normal and emergency operations.

Learning Outcomes:
Intended outcomes for this course are listed in the following rubric. You will have ample opportunity and will be encouraged and assisted in developing additional personalized outcomes relevant to your own goals as a pilot. You may also find that you succeed in achieving some un-intended outcomes to take with you into all of your life and flying experiences.

Process Skills:
To effectively meet the performance tasks and objectives, you will need to learn and practice the following skills:

- Envisioning
- Critical thinking
- Collaboration
- Behavior modeling
- Problem solving
- Identify personal attributes
- Evaluation of self and others
- Goal setting
- Organize concepts
<table>
<thead>
<tr>
<th>LEARNER OUTCOMES</th>
<th>ASSESSMENT STRATEGIES</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Upon successful completion of this course</strong> Each student will be able to:</td>
<td>Each Student will:</td>
</tr>
</tbody>
</table>
| 1. Demonstrate an understanding of the factors related to why pilots prioritize tasks in the cockpit, including how those factors were determined and their relationship to standard operating procedures, checklists, and flow checks. | • Work in teams to study CTM theory and gather appropriate information about task prioritization factors, including how those factors related to practical application in the flight environment.  
• Participate in class discussions and answer homework and quiz questions. |
| 2. Identify and discuss various aspects of aviation human factors related to human error and cognitive processes related to human multitasking abilities. | • Participate in class discussions and collaborate to complete a short in class exercise on aviation human factors.  
• Answer questions from assigned readings, homework material, and quizzes. |
| 3. Demonstrate the ability to identify proper task prioritization and execution in various flight regimes, including normal and emergency operations. | • Present a short oral briefing to the class of their analysis of an aviation accident or incident in the context of CTM performance.  
• Participate in class discussions, answer questions from assigned readings, homework material, and quizzes. |
| 4. Demonstrate the necessary organizational and communications skills to apply their understanding of effective task prioritization to cockpit operations. | • Analyze scenarios taken from the flight environment with respect to proper and improper prioritization techniques and relate those techniques to safety of operations. |
| 5. Demonstrate their ability to recognize proper and improper CTM performance in their own flying. | • Write a short introspective assessment of their last two aircraft or simulators flights with respect to their ability to prioritize tasks as appropriate with regards to standard operating procedures, checklists, flow checks, and other factors related to effective CTM. |
**Performance Tasks:**

The following activities and tasks will build upon one another and provide you the opportunity to demonstrate and assess your mastery of the learning outcomes:

<table>
<thead>
<tr>
<th>Written tasks</th>
<th>Oral tasks</th>
</tr>
</thead>
<tbody>
<tr>
<td>You will keep a journal for a minimum of two airplane or simulator flights in which you reflect on your level of situational awareness, using specific examples of times when you had high or low situational awareness. Also, you will identify times of highest and lowest workload during those flights and comment on how well you were able to prioritize cockpit tasks, follow proper standard operating procedures, checklists, and flow checks during those times of high and low workload. The assignment will include a short synopsis of your notes relating your performance to what you have learned in this course about how to effectively prioritize tasks in the cockpit.</td>
<td>You will also complete a short homework assignment analyzing a scenario from either an NTSB or ASRS report and describe proper or improper prioritization techniques and related to that scenario with respect to safety of operations. You will complete two quizzes that examine your understanding of course content and objectives.</td>
</tr>
</tbody>
</table>

**Oral tasks**

You will participate in class discussions, group projects, and present a short oral briefing to the class of their analysis of an aviation accident or incident in the context of CTM performance.

**Learning Resources:**

I will provide learning resources in the form of handouts and web sites for reference. Also, you will be asked to investigate and share with your peers at least one other resource, including books, periodicals, personal contacts, web sites, or others that can assist your understanding and application of the concepts and issues addressed in this course.
Concepts, Themes, and Issues:

During this course you will need to gain a working understanding of many concepts, themes, and issues to facilitate your insight into the content and your application of knowledge, skills, and practices. The following are a basis on which you should build your own comprehensive list:

**Concepts**
- Aviation human factors
- Human error
- Situational awareness
- Attention management
- Cognitive processes
- Concurrent task management

**Themes**
- Task prioritization
- Goal setting
- Effective communications
- Building confidence

**Issues**
- Multitasking abilities
- Assessment; self-assessment
- Building confidence

Instructor's Expectations and Assessments:

I expect you to participate fully in class discussions. You will assess your own performance based on your mastery of the concepts and ability to improve your CTM performance in the cockpit. In class activities and quizzes are designed to help you with that assessment. Your performance does not reflect my opinion of you as a person or your worth as an individual.

*Above all: immerse yourself in the learning experience and enjoy the process*

*Have Fun!!*

ADA Statement:

*Students who have special needs or disabilities that may affect their ability to access information and/or material presented in this course are encouraged to contact me or Robert Campbell, ADA Compliance Officer, Director, ADA Affairs and Student Assistance on campus at 963-2171 for additional disability-related educational accommodations.*
APPENDIX C
LINE ORIENTED FLIGHT TRAINING (LOFT) PROFILE USED FOR PRETEST AND POSTTEST FTD FLIGHTS

All simulated flights began at Renton Airport in Renton, Washington. Pilots briefed procedures prior to their flight as they would in a normal flight planning situation. After takeoff all pilots flew a radar vector departure and contacted Seattle Center for vectors to the instrument landing system (ILS) approach to Runway 13R at Boeing Field in Seattle. After flying the missed approach procedure (MAP) pilots held at BLAKO intersection as published and after entering the hold were given directions by Air Traffic Control (ATC) to fly pilot navigation direct to NEEAL intersection for the ILS Runway 34R at Seattle-Tacoma International Airport in Seattle. At a distance of approximately 15 miles southeast of Seattle, ATC gave the pilots radar vectors for the localizer course intercept for the ILS Runway 34R.

Radio frequencies and procedures adhered to current published data. Altitudes, vectors, and intercepts followed standard MSA altitudes as depicted either on NACA approach plates, minimum enroute altitudes as depicted on enroute charts, or Seattle Air Route Traffic Control Center designated minimum vectoring altitudes as described in the Seattle ARTCC Standard Operating Procedures Manual.

All ATC communications followed standard phraseology as detailed by FAA Order 7110.65P and the FAA Aeronautical Information Manual Pilot/Controller Glossary. Additionally, all transfer of control within and between ATC sectors and areas followed standard procedure as dictated by FAA Order 7110.65P. The IAP charts for BFI and SEA are shown on the following pages.
SEATTLE, WASHINGTON

LOC/DME I-SEA

110.3
Ch 40
APP CRS
341°
Rwy 34R IDg
19901
TDZE
372
Apt Elev
433
Rwy 34L IDg
387
TDZE
433
MALS R
Rwy 34L/R

ILS RWY 34R
SEATTLE-TACOMA INTL (SEA)

ATIS
SEATTLE APP CON
118.0
SEATTLE TOWER
119.9
GND CON
121.7
CNC DEL
128.0

DME REQUIRED

VECTOR FACILITIES

R-341
DOOVE
SEA
3.3
(4F)

NEEAL
I-SEA
6.5

LOCALIZER 110.3
I-SEA

SILS1EP
341
760/50
(400-U
760-fl'

CIRCLING
1000-1
567 (600-1)
1000-1½
567 (600-1½)
1000-2
567 (600-2)

CATEGORY
A
B
C
D
S-LOC 34R
720/24
348 (300-½)
348 (300-½)
SIDESTEP 34L
760/50
373 (400-½)
373 (400-½)
CIRCLING
1000-1
567 (600-1)
1000-1½
567 (600-1½)
1000-2
567 (600-2)

Orig-0 05020

Retrieved from http://204.108.4.16/d-tpp/0501/00582134R.PDF
Following is the outline for ATC/pilot procedures and communications. The original LOFT used for the study was broken into sections based on the flight segment and challenge point. The LOFT here is not presented the same way due to formatting requirements for this document; the content is unchanged. Because it would not be possible to anticipate every possible pilot response, communications relied on the expertise of the instructor conducting the study to maintain proper phraseology and procedure as outlined in the pilot/controller glossary and according to standard ATC operations procedures. Task prioritization challenge points used in LOFT are also shown, and instructors that conducted the LOFTS adhered to the script as much as possible with only minor variations when necessary, but did not interfere with the CTM challenges.

All errors related to altitudes, headings, bank angles, course intercepts, procedure turns, VDPs, DAs, MDAs, MAPs, holds, procedure turns, ATC communications and timed approaches will be based on completion standards as stated in FAA-S-8081-4C Instrument PTS.

Number of challenge points: 14
Total errors: 20
A/N errors: 7
A/C errors: 7
N/C errors: 5
A/A: 1
CTM LOFT
Renton airport, Radar Vector Departure, Boeing Field airport, Missed Approach, hold at Blako, Radar Vectors to SeaTac

<table>
<thead>
<tr>
<th>Event</th>
<th>Pilot action</th>
<th>Challenge</th>
<th>Description</th>
<th>Type</th>
<th>ATC communication</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pre Takeoff</td>
<td>Configure instruments and systems</td>
<td>Call for IFR clearance</td>
<td>Errors are classified as:</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Read back clearance and clarify</td>
<td>A/N = aviate/navigate</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>A/C = aviate/communicate</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>N/C = navigate/communicate</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>A/A = aviate/aviate</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Takeoff</td>
<td>Call for takeoff clearance and read back</td>
<td></td>
<td>Amend vector for traffic in the middle of climb checklist (A/N error if pilot fails to complete climb checklist prior to turn and A/C if pilot fails to complete climb check prior to calling ATC)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Climb</td>
<td>Accomplish climb checklist</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Read back and turn 350</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Renton ATIS (126.95):** Renton airport information Delta, 2253 Zulu, wind 180 at 3, visibility 4, mist, 800 overcast, temperature 11, dewpoint 5, altimeter 30.15. Renton landing and departing runway 15. Expect an NDB approach runway 15. Advise the controller on initial contact you have ATIS information Delta.

**Renton Ground (121.6):** Frasca 141 cleared to Boeing Field International airport via radar vectors. Climb and maintain 3000, contact Seattle approach 119.2 when advised. Squawk 3135.

**Renton Ground:** Read back correct (or modify)

**Renton Tower (124.7):** Frasca 141, cleared for takeoff runway 15 fly runway heading to 1000 then left turn heading 310 radar vector to Boeing Field

**Renton Tower:** Frasca 141 fly heading 350 vectors for traffic
<table>
<thead>
<tr>
<th>Level off and cruise</th>
<th>Direct pilot to contact approach 100 feet prior to level off (A/C error if pilot calls ATC and as a result deviates from altitude)</th>
<th>Renton Tower: Frasca 141 Contact Seattle Approach 119.2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Accomplish cruise checklist</td>
<td>A/C</td>
</tr>
<tr>
<td></td>
<td>Follow and read back or clarify ATC instructions</td>
<td>Boeing ATIS (127.75): Boeing Field international airport information Whiskey, 2250 Zulu, wind 130 at 4, visibility 2, light rain. Temperature 10, dewpoint 3, altimeter 30.22. Boeing is landing and departing runways 13 right and left. Expect an ILS runway 13 right. Advise the controller on initial contact you have information Whiskey.</td>
</tr>
<tr>
<td></td>
<td>Give vector and frequency change in the middle of approach briefing (A/N error if pilot fails to complete approach checklist before turning and A/C if fail to complete it before calling ATC)</td>
<td>A/N A/C</td>
</tr>
<tr>
<td>Intermediate approach segment</td>
<td>Follow and read back or clarify ATC instructions</td>
<td>Seattle Approach: Frasca 141, turn left heading 310, contact Seattle Approach on 123.4</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Boeing ATIS (127.75): Boeing Field international airport information Whiskey, 2250 Zulu, wind 130 at 4, visibility 2, light rain. Temperature 10, dewpoint 3, altimeter 30.22. Boeing is landing and departing runways 13 right and left. Expect an ILS runway 13 right. Advise the controller on initial contact you have information Whiskey.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>NOTE: try for intercept approximately 2-3 miles outside the marker</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Level of difficulty</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Level of difficulty</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Level of difficulty</th>
</tr>
</thead>
<tbody>
<tr>
<td>A/C</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Level of difficulty</th>
</tr>
</thead>
<tbody>
<tr>
<td>A/N A/C</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Level of difficulty</th>
</tr>
</thead>
<tbody>
<tr>
<td>Seattle Approach</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Level of difficulty</th>
</tr>
</thead>
<tbody>
<tr>
<td>Boeing ATIS</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Level of difficulty</th>
</tr>
</thead>
<tbody>
<tr>
<td>Seattle Approach</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Level of difficulty</th>
</tr>
</thead>
<tbody>
<tr>
<td>NOTE</td>
</tr>
<tr>
<td>Final approach segment</td>
</tr>
<tr>
<td>------------------------</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Five Ts at the marker (Contact tower)</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Missed approach segment</td>
</tr>
<tr>
<td>-------------------------</td>
</tr>
<tr>
<td>Execute go-around</td>
</tr>
<tr>
<td>Accomplish climb checklist</td>
</tr>
<tr>
<td>Contact departure</td>
</tr>
<tr>
<td>Read back ATC instructions</td>
</tr>
<tr>
<td>7</td>
</tr>
<tr>
<td>8</td>
</tr>
<tr>
<td>Fail localizer just before MAP (A/N if pilot deviates from proper descent or airspeed <em>due to</em> dealing with localizer failure)</td>
</tr>
<tr>
<td>Give go around due to ILS out of service (A/C if pilot calls ATC before initiating MAP or before completing climb checklist)</td>
</tr>
<tr>
<td>Put ILS operational</td>
</tr>
<tr>
<td>Monitor Missed approach regarding altitude changes and identification of Cogar (A/N if altitude or airspeed deviate <em>due to</em> pilot identifying Cogar)</td>
</tr>
<tr>
<td>Boeing Tower: Frasca 141 ILS out of service go missed approach <em>[if pilot fails to execute MAP as per approach plate strip-otherwise respond]</em>] and contact Seattle Approach 119.2</td>
</tr>
<tr>
<td>Seattle Approach (119.2): Frasca 141, radar contact [#] miles from Boeing Field, ILS now operational. Execute the published missed approach, hold Southeast of Blako as published, maintain 5000.</td>
</tr>
<tr>
<td>Enroute and hold</td>
</tr>
<tr>
<td>------------------</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Leaving hold and descent</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Segment</td>
</tr>
<tr>
<td>---------</td>
</tr>
<tr>
<td>Initial Approach</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Intermediate Approach</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Final Approach</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Landing</td>
</tr>
</tbody>
</table>
The purpose of the pilot study was to refine techniques to be used in the main study, to establish a working definition of CTM errors, and to investigate issues related to experimental design and validity. The pilot study was conducted six months prior to the main experiments with approval from CWU Human Subjects Review Committee and Oregon State University Institutional Review Board.

The pilot study had 16 participants enrolled as degree-seeking students in the CWU Flight Technology Department who were all at the same stage of training. An informed consent document (Appendix A) was presented and explained to each participant by the investigator before they consented to participate in the study. Design, delivery, and discussion of data collected during the pilot study are summarized here.

Design and Delivery

The experimental procedure was conducted with two groups (a control group and an experimental group) comprised of eight participants each. For the observation portions of the experiment, each participant in each group flew one-hour pretest flight and a one-hour posttest flight on the CWU Frasca 242 FTD, which simulates a normally aspirated conventional twin engine with full feathering propellers and accumulators, dual boost pumps, and dual vacuum pumps. The Avionics package is a Bendix/King stack including dual KY 196 Coms, dual KN 53 nav radios, KDI 572 DME, KNS 81 RNAV, KR 87 ADF, KT76A transponder, KFC 150 Flight director, KMA 24 audio panel with marker beacons, and KC 192 Autopilot. The Frasca 242 is equipped with EFIS, including EADI (electronic attitude direction indicator) and EHSI (electronic horizontal situation indicator).

Pilots flew their pretest and posttest observation flights approximately two weeks apart. Participants operated the Frasca 242 as they do in their normal flight
training. All assigned flight tasks followed standard operating procedures as designated by the Federal Aviation Administration (FAA) and the CWU standard course of training.

FTD flights were conducted in a LOFT format using standard Instrument charts and procedures. Flights began at Snohomish County Airport, Paine Field, in Everett Washington. Pilots flew the Paine Two departure and contacted Seattle Center for vectors to the instrument landing system (ILS) approach to Runway 13R at Boeing Field in Seattle. After flying the missed approach procedure (MAP) pilots held at BLAKO intersection until given directions by Air Traffic Control (ATC) to fly pilot navigation to ANVIL intersection for the ILS Runway 16R at Seattle-Tacoma International Airport in Seattle. At a distance of 15 miles southeast of Anvil, pilots were switched to radar vectors for the localizer course intercept for the ILS Runway 16R. This particular LOFT was chosen because it involves many complicated procedures in a short time frame and the pilots were in a high workload environment for most of the one hour flight. The LOFT had to be modified for the main study because the Frasca 242 simulates a faster airplane than the Frasca 141 and the entire LOFT would have taken much longer.

CTM errors were observed and classified according to the ANC hierarchy. Instead of using discrete challenge points, flights and videotapes were observed and analyzed to classify and record all CTM errors committed during the flight. The definition of a performance error was based on the FAA Practical Test Standards stated by FAA-S-8081-4C with respect to altitude, airspeed, heading, intercepting and tracking course, use of checklists, procedures, and ATC communications. Many of the challenge points defined in Appendix C, for example numbers 2, 3, 4, 7, and 8 were based on errors made in the pilot study.

Another important outcome from the pilot study was that it gave the investigator a chance to evaluate conduct of the CTM short course and make changes as appropriate for the main study.