2012

The Utility Of Verbal Display Redundancy In Managing Pilot's Cognitive Load During Controller-pilot Voice Communications

Daniela Kratchounova
University of Central Florida

STARS Citation

Kratchounova, Daniela, "The Utility Of Verbal Display Redundancy In Managing Pilot's Cognitive Load During Controller-pilot Voice Communications" (2012). *Electronic Theses and Dissertations, 2004-2019*. 2317.
https://stars.library.ucf.edu/etd/2317

THE UTILITY OF VERBAL DISPLAY REDUNDANCY IN MANAGING PILOT'S COGNITIVE LOAD DURING CONTROLLER-PILOT VOICE COMMUNICATIONS

by

DANIELA KRATCHOUNOVA
M.S. Embry-Riddle Aeronautical University, 1999
M.S. Technical University of Sofia, 1986

A dissertation submitted in partial fulfillment of the requirements
for Degree of Doctor of Philosophy
in the Department of Psychology, Applied Experimental Human Factors
in the College of Sciences
at the University of Central Florida
Orlando, Florida

Spring Term
2012

Major Professor: Florian Jentsch

ABSTRACT

Miscommunication between controllers and pilots, potentially resulting from a high pilot cognitive load, has been a causal or contributing factor in a large number of aviation accidents. In this context, failure to communicate can be attributed, among other factors, to an inadequate human-system interface design, the related high cognitive load imposed on the pilot, and poor performance reflected by a higher error rate. To date, voice radio remains in service without any means for managing pilot cognitive load by design (as opposed to training or procedures). Such an oversight is what prompted this dissertation. The goals of this study were (a) to investigate the utility of a voice-to-text transcription (V-T-T) of ATC clearances in managing pilot's cognitive load during controller-pilot communications within the context of a modern flight deck environment, and (b) to validate whether a model of variable relationships which is generated in the domain of learning and instruction would "transfer", and to what extend, to an operational domain. First, within the theoretical framework built for this dissertation, all the pertaining factors were analyzed. Second, by using the process of synthesis, and based on guidelines generated from that theoretical framework, a redundant verbal display of ATC clearances (i.e., a V-T-T) was constructed. Third, the synthesized device was empirically examined. Thirty four pilots participated in the study – seventeen pilots with 100-250 total flight hours and seventeen with >500 total flight hours. All participants had flown within sixty days prior to attending the study. The experiment was conducted one pilot at a time in 2.5-hour blocks. A 2 Verbal Display Redundancy (no-redundancy and redundancy) X 2 Verbal Input Complexity (low and high) X 2 Level of Expertise (novices and experts) mixed-model design was used for the study with 5 IFR clearances in each Redundancy X Complexity condition. The results showed that the amounts of reduction of cognitive load and improvement of performance, when verbal display redundancy was provided, were in the range of about 20%. These results indicated that V-T-T is a device which has a tremendous potential to serve as (a) a pilot memory aid, (b) a way to verify a clearance has been captured correctly without having to make a “Say again” call, and (c) to ultimately improve the margin of safety by reducing the propensity for human error for the majority of pilot populations including those with English as a second language. Fourth, the results from the validation of theoretical models “transfer” showed that although cognitive load remained as a significant predictor of performance, both complexity and redundancy also had unique significant effects on performance. Furthermore, these results indicated that the relationship between these variables was not as “clear-cut” in the operational domain investigated here as the models from the domain of learning and instruction suggested. Until further research is conducted, (a) to investigate how changes in the operational task settings via adding additional coding (e.g., permanent record of clearances which can serve as both a memory aid and a way to verify a clearance is captured correctly) affect performance through mechanisms other than cognitive load; and (b) unless the theoretical models are modified to reflect how changes in the input variables impact the outcome in a variety of ways; a degree of prudence should be exercised when the results from the model “transfer” validation are applied to operational environments similar to the one investigated in this dissertation research.

To my kids, Elitza and Kaloyan

ACKNOWLEDGMENTS

The support and encouragement of a large cast of friends and family, friends who are like family to me, mentors, and colleagues helped me complete this amazing journey. This all-star cast spans over four continents, at least five different native languages, and a very diverse set of cultures and origins. I will introduce the cast chronologically.

At the beginning, there were Dr. Wise – the best mentor (period), and Eric Vaden – my best friend and the brother I never had. During my years at Embry-Riddle Aeronautical University (ERAU), Dr. Wise used to say, I know you would love to teach one day and in order to do that you need a PhD. The reader would have to exercise some patience while following the chronology here but one day in the summer of 1999 I got a call at work from Dr. Wise's office and the voice on the other side spoke perfect Bulgarian. This is how I met Dr. Mouloua who later was one of my professors at UCF, and now is a member of my dissertation committee.

Following the events of September 11, 2001 I lost my job. Now is the time to start your PhD, said Dr. Wise to me then, because the economy is down. In a couple of years when it picks back up, he continued, you will have become even more marketable than you are today. Interestingly enough, a very similar argument brought me to this great country, and you guessed it, it was Dr. Wise who said that to me just a few years earlier when I first met him while visiting the United States from Bulgaria. I came to the US as a gradate student in the spring of 1998, and since then there has not been a single important decision about my life or kids, that I have made, which has not been directly influenced by Dr. Wise and Eric.

At the beginning of December 2001, Dr. Wise and I drove to Orlando to visit with the UCF Department of Psychology. He wanted to introduce me to the faculty there as a prospective

PhD student. Just short of one year later, I met Dr. Florian Jentsch and Dr. Peter Hancock. They have imparted a great deal of wisdom and knowledge on me. Dr. Jentsch's boundless patience as my academic adviser paired with his hallmark pedagogy taught me many lessons but the following I will remember for the rest of my life:

- First, I should never spread myself too thin in trying to accomplish too many things at the same time, because my performance would inevitably suffer on all of them. If I were to add something to that maxim, I would say that there is no such thing as multitasking, and one can only try to multiplex, where the number of "slices" one can "cut" time into, is finite
- Second, it takes time for a dissertation idea to grow and take shape. My addition – this applies to everything that is worthwhile
- Third, a PhD is a "thinking" degree. One should be able to take time to think, not only do, do, and do. Tell me about it!

My deepest gratitude goes to all four members of my committee - Dr. Jentsch (Chair), Dr. Hancock, Dr. Mouloua, and Dr. Wise - for helping me not become a statistic on the side of the 80% of graduate students who never complete their degree if they get a job before finishing it. Starting to get the picture? Wait, I am not even halfway through the cast.

The period between 2005 and 2010 can be labeled as "do, do, and do" and long hours of it. Many of the email reminders from Dr. Wise to finish my degree were left unanswered. The truth is that it is hard to switch to thinking mode when your brain is "fried" after 50-60 hours of do, do, and more do, every week.

In 2010, I had the honor of meeting one of the two best safety pilots in the Universe, the one who owns “that” Universe’s Eastern hemisphere – Hen Raz. He will never give himself credit for any of the following, but I am giving the credits here, right? In early 2011, he did the pilot study for this dissertation. His incredible insight and ability to dissect with a surgical precision, and almost instantaneously find the “holes” in just about anything he cares to look at, along with his genuine enthusiasm about Human Factors, is what inspired me to put everything back in high gear. Hen’s friendship and camaraderie spans over 7,000 miles but those 7, 000 miles mean nothing when I need his help. After a 30-hour overseas trip with almost no sleep, he was here for me during the home stretch of my data collection (I will come back to the data collection later in this chronology) in the fall of 2011. His complete lack of pretence, quiet resolve, ingenuous demeanor, and unconditional support for all my endeavors are nothing short of remarkable.

The Western half of the Universe of safety pilots is reigned by Bob Wilson. I owe Bob, and his outstanding subject matter expertise, the very sophisticated clearance read-back scoring system used for this dissertation. For those readers who are wondering about the whole safety pilots and Universe thing, these are the two people who, I know, will catch me if I were to fall backwards without looking. While used here as a metaphor, this makes for a great team-building exercise. Please proceed with caution! It takes this level of trust in order to achieve great team performance, and Bob and Hen demonstrated that brilliantly, over and over again, during the course of 2011.

My dissertation study required some pretty complex programming within the X-Plane ® simulation in order to be even remotely successful. I shared my concerns with Ellie, my daughter, and she said, “Mom, you got to find yourself a geek”. Without going into of the domain of linguistics, I would not be able to explain why this cannot be translated from Bulgarian into English with the same amount of humor in it. You will have to trust me that I was in tears laughing when she said it. How do you find yourself a geek? And more importantly, how do you know if he or she is “your” geek? It is not easy and you will know it when you found him or her. My geek’s name is Chris Nusbaum and he deserves every bit of the title – The Best Programmer Ever.

In the fall of 2011, I was determined to complete my data collection, and not only finish it, but finish it within a week – all that was left from my vacation days from work. Eric, the brother I mentioned at the beginning, flat out said that I was crazy to even have a hint of a trace of a shadow of a suggestion of a thought about it being possible. Of course he was joking. He launched a recruiting campaign of epic proportions and with the help of the absolutely wonderful faculty, staff, and students (especial thanks go to Shannon Cummings and Julian Archer) at ERAU’s Department of Human Factors and Systems (HFS); I managed to complete my data collection in a week and two weekends. I call the HFS department - my “home department” - and I mean it from the bottom of my heart.

The support and encouragement of my kids, Elitza (a.k.a. Ellie) and Kaloyan (a.k.a. K), to complete my degree has been parent-like for all these years. I dedicate this dissertation to them.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER ONE: INTRODUCTION

Air Traffic Management (ATM) is a complex system, which is heavily reliant on the sub-system of voice radio communications to support a comprehensive information transfer. Today, virtually all information exchanges between pilots and controllers are carried by voice. Historic analysis of aviation accidents, however, has identified that a breakdown in effective human communication has been a causal or contributing factor in the majority of accidents (Wiegmann & Shappell, 2001). Two groups of potential sources of communications errors involving the voice radio communication system could be identified in the reports listed in the Aviation Safety Reporting System (ASRS) database: (a) the operational environment, and (b) the pilot-system interface design. For example, the reduced intelligibility of voice communications due to interference and noise is a well-known operational environment attribute leading to a higher likelihood for misinterpretations and miscommunications, simultaneous transmissions, call sign confusion, and read-back / hear-back problems (Kerns, 1991). Furthermore, some pilot-system interface features, where information is presented without carefully considering the capabilities and limitations of human information processing, and specifically of working memory, have the potential for inducing significant levels of cognitive load, and creating a flight deck environment prone to error. For instance, visual information displays on modern flight decks include significant levels of redundancy provided via interactive graphical user interfaces and information is presented in a multimedia fashion. However, displays of auditory verbal information (historically plagued with distortion, interference, and noise) lack the ability to support pilot’s working memory and still rely on training and procedures to ensure information transfer.

Over time, to mitigate the risk of communication errors, a significant effort has been made by the aviation community to upgrade voice radio communications equipment, improve operational procedures, and radio discipline. For example, the use of English as a shared code, standard phraseology, and a voice communications protocol were mandated. By following a standard phraseology format, the task to assemble information into a message was made quick and straightforward, and the procedures to support message exchange - consistent for all messages (Flathers II, 1987). In addition, all controller-pilot communication transactions require an acknowledgement in order to substantiate the reception of the correct information by the correct party. Yet, in the context of the compound nature of the problem, the intended benefits of these mostly procedural (as opposed to system design) solutions are frequently overcome by the drawbacks (Kerns, 1991). More specifically, the very nature of voice communications does not allow either the pilot or the controller to defer the handling and disposition of incoming messages to another, perhaps less busy time. Regardless of the criticality of a particular message, the receiving party must attend to it immediately, and make the appropriate acknowledgement. As the retention and assurance of message integrity, delivered through language, is of a critical importance to the entire ATM system, to send or receive a single message, pilots and controllers are required to perform many “overhead” tasks, including monitoring the voice channel, filtering voice channel traffic for his/her own aircraft identifier, acknowledging the receipt, etc (Flathers II, 1987). Although these tasks are put in place to ensure the integrity information transfer, they also generate a significant amount of workload for the pilot and the controller (Rehmann, 1995; 1997). In addition, a poor microphone technique, a rapid rate of speech delivery, use of non-standard phraseology, and accents may further reduce the effectiveness of voice radio communications and negatively impact pilots' cognitive performance.

Today, pilots learn to shorthand the clearances they receive from the ATC and do so by using a specific order or format (e.g., cleared to, altitude, route, and frequency). Although there is no requirement that an entire clearance be read back, pilots are expected to read back the parts of any clearance containing altitude assignments, radar vectors, or any portion of the clearance requiring verification. ATC may request a clearance read-back when certain factors such as clearance complexity suggest a need. However, under certain conditions (e.g., a high workload phase of flight, a signal with poor intelligibility, or less experienced crew) in single-pilot operations, writing down clearances could impose increased working memory load, and temporarily distract the pilot from the primary tasks of aviating and navigating the aircraft. Although less likely, in multi-crew operations where tasks are divided between the pilot flying (PF) who is responsible for the primary tasks of aviating and navigating the aircraft, and the pilot monitoring (PM) who is responsible for communicating with the air-traffic control, monitoring systems, etc., the effects may be very similar in nature.

For over 20 years now, the aviation community has viewed a partial replacement of voice radio with data link communication technology as the ultimate means to address the limitations of voice radio while preserving its positive aspects. This solution is a globally coordinated effort of local and international aviation authorities in conjunction with airlines and avionics manufacturers, which to this day, is a work in progress. While the potential for relieving some of the frequency congestion, and generally enhancing air-to-ground communications for appropriately equipped aircraft certainly exists, this is only a palliative solution because even after the implementation of data link technology, all time-critical communications will continue to be conducted by voice.

This dissertation addresses the specific challenges associated with the remaining role (outside the scope of data link communications) of controller-pilot voice communications by proposing a solution, which provides the pilots with a redundant means to access, remember, and verify the content of controller's messages. More specifically, the solution consists of a redundant display of the clearances received via voice in a form of text. This automatic voice-to-text (V-T-T) transcription of the controller's message would be available for viewing immediately after it is delivered by voice, as well as, at any time during the flight at pilot's discretion. The utility of such solution, as a means to reduce the potential for elevated cognitive load and communication errors, will depend on factors associated with: (a) the human information processing system, (b) the task, and the task environment; (c) the inherent attributes of the system's operational environment, and (d) the individual differences in the users' population.

Problem Statement

Voice radio remains in service without means for managing pilot cognitive load, during controller-pilot voice communications, by design. Thanks to the remarkable progress made in electronic display technology in recent years, the presentation of visual information in modern flight decks includes significant levels of redundancy. Yet, the display of auditory verbal information (e.g., controller-pilot voice communications) had been overlooked in that regard. To this day, the integrity of information transfer via voice radio communications relies on the pilot to correctly capture, remember, and act upon the controller's message by following certain rules and operating procedures. Nevertheless, even after the implementation of data link communications technology, voice radio will remain as the primary means of obtaining time-critical and tactical agreements, which address local events or short-term conditions (Flathers II, 1987). Generated throughout a flight, voice communications will continue to fill in the unspecified details of strategic agreements already established via data link. However, along with all beneficial features controller-pilot voice communications have to offer (i.e., “party line”, practically unlimited flexibility, always “in–the-loop”, etc.), its well-known limitations, such as (a) the inability to defer the handling and disposition of incoming messages; (b) the many “overhead” tasks potentially generating significant amount of workload for the pilot and the controller; or (c) the poor intelligibility due to interference and noise, will continue to exist even within this somewhat limited, yet, critical for the safety of flight deployment. Conceivably, an added layer of redundancy to the display of such important type of information could mitigate the harmful effects of these drawbacks to a large extent by: (a) supporting pilots’ information processing, and therefore reducing the potential for cognitive load and communication errors; and (b) maintaining the integrity of information transfer, not only during the period between the time a message is delivered, and the time it is acted upon but also at any time during a flight at pilot’s discretion without changing the long established rules and procedures in aviation.

This dissertation addressed the underlining mechanisms in working memory responsible for the processing of verbal information, the factors, and conditions influencing that processing; and was focused on the investigation of the utility of such redundant display in minimizing the potential for increased pilot cognitive load. The investigation was conducted in the context of modern flight deck operations, which generally include a multimedia type of information presentation that requires the pilots to engage in multimodal interactions with highly integrated complex system environment. The potential for success of the solution proposed here was determined by answering the following question:

- As compared to the use of voice alone (i.e., without added redundancy), and in the context of a modern flight deck environment, what is the utility of a voice-to-text transcription of controller's messages in managing pilot's cognitive load?

The Remaining Role of Voice Communications in the Future ATM Environment

In the U.S., the Federal Aviation Administration (FAA) has mapped a plan for building the future air traffic management (ATM) system which will use advanced communications, navigation, and surveillance technologies to support global flight planning, aircraft operation, and air traffic control (ATC) services. One of the important components of this ATM system will be the use of Controller-Pilot Data Link Communication (CPDLC) to transmit ATC clearance information between air traffic controllers and flight crews. CPDLC systems are a digital means to transmit information between the air traffic controller and the pilot. The use of CPDLC in high-density airspace offers the potential to relieve some of the frequency congestion, enhancing existing air-ground communications, and offering unambiguous transmission of routine and/or strategic messages between controllers and pilots. CPDLC will enable controllers to transmit text-based strategic ATC communications to appropriately equipped aircraft. This information will be presented on a display in the cockpit in a form of a text message. Pilots will also be able to acknowledge the receipt of that information, as well as transmit requests to air traffic controllers by the use of standardized, pre-formatted text messages.

While the implementation of CPDLC do have the potential to help overcome some of the drawbacks of voice communications by replacing voice as primary means of delivery for strategic controller-pilot communication, and to significantly improve the overall safety of the ATM system, it is nevertheless only a partial solution. CPDLC is not intended to fully replace voice radio communications. Rather, pilots and air traffic controllers will work within a dual voice/data link communication environment and choose whatever means are the most appropriate at the time. If the exchange is not time-critical, then they will have the choice of using voice or data link depending on operational circumstances. If tactical and time-critical communications are required, the controllers and crews will, as they do today, continue to use voice.

Current voice communication procedures require pilots and controllers to always be "in-the-loop" of information exchange. This allows both parties to determine the relevance of the exchanged information while maintaining constant awareness of the status of the entire communication system (e.g., procedures, equipment, and communication partner). The practically unlimited flexibility of spoken language (e.g., voice intonation and inflexion) enables pilot and controller to reach understanding in a variety of situations (even in the context of standard phraseology). The current voice radio communication environment also allows processes, such as:

- Negotiating an ATC clearance
- Obtaining knowledge of events and conditions that might affect the flight (e.g., delays, weather, and traffic congestion) by listening to ATC communications with other aircraft (i.e., "party line" information).

These desirable features of voice radio communications will, without a doubt, be preserved in the evolving ATM system. For example, voice radio will continue to: (a) be used for tactical, and time-critical communications; (b) allow pilot and controller to quickly reach understanding in unusual situations, (c) let both parties to communicate instantaneously, and (d) provide a means to negotiate an ATC clearance. In order to maintain a continuous awareness of the communication environment, as before, pilot and controller will remain directly involved in the generation or receipt of a voice message. Using the "party line" will continue to help pilots develop an accurate mental model of their immediate environment, as well, as avoid any adverse situations (Rehmann, 1997).

Voice communication tasks carried by a pilot usually involve receiving, processing, and acting upon instructions issued by a controller therefore requiring the pilot to retain the information in his/her memory for a short period of time between the receiving and acting upon these instructions (Loftus, Dark & Williams, 1979). In addition, working memory load may be generated if, among other factors: (a) the controller's messages are longer, and more complex, containing several interrelated instructions; (b) the usability of the pilot-system interface is inadequate, (c) the controller's messages are with poor intelligibility, (d) it is a high-workload phase of flight, (e) the pilot lacks experience, and (f) the pilot is engaged in other tasks such as consulting a map, a checklist, or a chart (Loftus, Dark & Williams, 1979).

In summary, the outstanding challenge remains – while taking into account both, the operational environment (i.e., all aspects of the remaining role of voice communications), and the capabilities and limitations of human information processing system, augment the verbal display design such that the potential of higher cognitive load is minimized.

The Concept of Display Redundancy in Aviation

The term redundancy as applied to aviation displays was not introduced in the literature until it became apparent that the successful implementation and operation of data link communication technologies was largely dependent on how the respective human factors issues (e.g., pilot-system interface) were addressed. Yet, more than 20 years later, no consensus exists among the researchers in the aviation community regarding the benefits of redundant displays in the context of controller-pilot communication systems ((McGann, Morrow, Rodvold & Mackintosh, 1998; Farley, Hansman, Amonlirdviman & Endsley, 2000; Helleberg & Wickens, 2003; Wickens, Goh, Helleberg & Talleur, 2002). The majority of human information processing models employed in the research of the effects of data link pilot-system interface modality on pilot performance “predict” that due to different processing resources associated with visual and auditory modalities redundant displays may clearly provide the “best of both worlds” (Helleberg & Wickens, 2003, p. 193). However, their investigation of the effects of simultaneous redundant display of data link in a context of a multiple-task environment typical for single-pilot operations showed that display redundancy not only presented many of the same benefits as the visual display alone but also in some cases was inferior to the visual-only display. The investigation did not go outside the paradigm of the association of different processing resources with visual and auditory modalities, and it stopped short of addressing the processing visual and auditory verbal information within working memory. In contrast, Lancaster and Casali (2005) found that for single-pilot general aviation operations: (a) a textual controller-pilot communication presentation increased response time and workload, (b) it was not desirable without a speech component, and (c) a redundant presentation was preferable to voice alone.

In summary, these seemingly contradicting results may be indicative of the lack of: (a) a clear discrimination in the employed theoretical frameworks between perception and processing of visual (text), and auditory (voice) verbal input; (b) a differentiation between the two separate sensory mechanisms involved in the perception of verbal input (i.e., visual and auditory), and the single working memory faculty (i.e., the Phonological loop) involved in its processing; and (c) an explicit identification of the role voice communications have in support of the mostly visual-spatial task of piloting an aircraft. Most importantly, the effects of verbal redundancy were examined only in the context of text (as opposed to voice) as the primary mode of a message delivery.

Scope of Research

First, a conceptual framework is introduced. Second, a comprehensive review of the related literature is presented focusing on the mechanisms within the working memory responsible for carrying out the processing of visual and auditory verbal information. Also, the factors and conditions that are thought to impact such processing, and could influence the utility of redundant verbal display, are isolated. Third, all these factors are categorized into: (a) information presentation design factors, and (b) individual differences factors; that can be manipulated, co-varied out, or fixed for the purposes of this dissertation. Ultimately, a subset of these factors, leading to this dissertation's research, is defined based on level of importance, practicality, and interest. This process of factor identification and categorization is believed to be essential for the creation of a very comprehensive portrayal of verbal display redundancy and the impact it has on cognitive load in a modern flight deck environment. The literature reviewed in the next section of this dissertation includes basic and applied research findings on:

1. The basic architecture of working memory, and the role of phonological loop as the faculty involved in the processing of visual and auditory verbal information
2. The sources of cognitive load
3. The impact of multimedia presentations on working memory processing
4. The pros and cons of utilizing display redundancy in general and verbal display redundancy in particular.

Each section of the review is followed by a summary of the factors impacting the processing of verbal information within working memory along with their relevance to this dissertation.

Conceptual Framework

The virtual absence of supporting empirical evidence for the utility of verbal display redundancy as a tool for managing pilot's cognitive load in the aviation literature warranted an inquiry into the research literature at large. The goal was to obtain theoretical support based on substantial empirical data, and a set of practical guidelines on how to present verbal information such that the outcome of controller-pilot voice communications is optimized within the voice communications system limitations, which will continue to exist as it is today, even after a wider implementation of digital data-link. The body of research to offer a wealth of empirical data including application guidelines was the instructional design literature in general, and the cognitive load and multimedia learning literature in particular. While aware of the potential challenges associated with the use of empirical findings from a learning domain into an operational domain (e.g., a flight deck) due to the inherently different nature of these environments, I based my decision to employ these guidelines on the following assumptions highlighting the common attributes associated with these two realms:

1. From a presentation of information standpoint, multimedia learning and a modern flight deck are two environments very rich in multimedia.
2. From a task performance standpoint, the execution of complex cognitive tasks such as learning or piloting requires real-time, active processing of new information within the working memory.

The conceptual framework introduced next provides a means to generate testable hypotheses about the utility of displaying a transcription of the controller-pilot voice communication (i.e., redundant display of the voice communications content), and a context where the interaction between working memory processes, the attributes of verbal display redundancy, and the characteristics of the environment their interaction takes place, can be researched and better understood.

Baddeley's Model of Working Memory

For its insight into the processing of visual and auditory verbal information within the phonological loop, and the way all working memory subsystems interact to support the loop when its limited capacity is reached, Baddeley's model of working memory (Baddeley, 1981; 1986; 1992; 1996; 1998; Baddeley & Hitch, 1994), and specifically the recently updated model of the phonological loop (Baddeley, 2003; Baddeley & Larsen, 2007), was selected as the basic research component of this dissertation's conceptual framework.

Cognitive Load Theory and Cognitive Theory of Multimedia Learning

The applied research component of this framework consists of two models: (a) the cognitive load theory (Sweller, 1988, 1994, Chandler & Sweller, 1991; Sweller, Van Merriënboer & Paas, 1998), and (b) the cognitive theory of multimedia learning (Mayer, 1997, 2001; 2005). Both models offer a wealth of empirical evidence on the impact of presentation modality and format (including redundant formats) on processing of novel information in working memory; as well as, a set of practical guidelines for the best way to present information such that the cognitive load imposed on the user is minimized. The two theories also share a set of common assumptions with all human-information processing (HIP) models including: (a) the critical role working memory plays in the performance of complex cognitive tasks, (b) its capacity limitations, and (c) the existence of separate memory resources for different input modalities. Furthermore, the two theories assert that information should be presented such that the limited working memory resources are used as efficiently as possible because cognitive overload can jeopardize learning outcomes particularly with multimedia instructions, where learners have to integrate different information sources like text, pictures, and narration (Chandler & Sweller, 1991; Mayer, Heiser & Lonn, 2001; Mayer & Sims, 1994). For example, the cognitive load imposed on working memory could become high when the integration of information presented in different modes requires an element presented in one mode to be held active in working memory while searching for the corresponding element in the other (Jeung, Chandler & Sweller, 1997; Leahy, Chandler & Sweller, 2003; Sweller & Chandler, 1994)). Working memory load can become even higher particularly when previous knowledge or experience is insufficient, and almost no schemata exist to steer the search process (Kalyuga, Chandler & Sweller, 1998). Therefore, when presenting information, the ultimate goal should be the prevention of cognitive overload by employing the limited working memory resources and modality-specific working memory systems, as optimally as possible (Kalyuga, Chandler & Sweller, 1999; Mousavi, Low & Sweller, 1995). It is essential to note that according to the cognitive load theory the limitations of working memory's capacity and duration apply only to the processing of new information acquired through sensory memory. Such limitations are non-existent when information is retrieved into working memory from long-term memory in a form of activated schemata (Pollock, Chandler & Sweller, 2002; Sweller, 2010).

In the context of this conceptual framework, I argue that the potential for a cognitive overload is a serious challenge not only in a multimedia learning environment but also in any environment where information is presented in a multimedia fashion and it supports a complex cognitive task which includes processing of new information. Furthermore, I make the assertion that after carefully accounting for the inherent attributes of an environment, techniques for managing cognitive load from a different research domain may be "borrowed" and successfully applied to the design of information displays as long as the underlying assumptions are the same. For example, a flight deck is by definition an operational environment that is very different as compare to a learning environment when it comes to: (a) the quality of the sensory input, and the visual and acoustic ambiance as a whole; (b) the ability to control the pace, timing, and length of information input; (c) the need for the user to attend other ongoing tasks, or (d) how time–critical or safety-critical is the task. However, it is reasonable to expect that if a particular approach to managing cognitive load has been deemed successful under comparable conditions in a learning environment (including information processing faculty, presentation format, and complexity of content) it has the potential to be successful in an operational environment, as well.

In summary, the most recently updated theoretical models of working memory (Baddeley, 2003; Baddeley & Larsen, 2007), cognitive load (Paas, Renkl, & Sweller, 2003; Sweller, van Merrienboer, & Paas, 1998; van Merriënboer & Sweller, 2005), and multimedia learning (Mayer, 1997, 2001; 2005) are employed as the underpinnings of this dissertation's conceptual framework. More specifically, this framework is utilized for the investigation of the utility of non-concurrent redundant verbal display (i.e., a transcript of pilot-controller voice communications in a form of text) in managing the pilot's cognitive load during pilot-controller voice communications through out a flight.

CHAPTER TWO: REVIEW OF THE RELATED LITERATURE

Working Memory

More than 25 years of research has recognized that working memory is comprised of multiple components associated with different modes, and it is responsible for information processing during the performance of complex cognitive tasks (Baddeley, 1981; 1986; 1992; 1996; 1998; Baddeley & Hitch, 1994). This limited capacity system allows the temporary storage and manipulation of information necessary for the performance of complex cognitive tasks such as comprehension, learning, and reasoning. According to Baddeley's model, working memory consists of three storage components, and an attentional control system. The storage components are: (a) the visuo-spatial sketchpad, (b) the phonological loop, and (c) the episodic buffer. The attentional control system is the central executive. Access to information from both long-term memory and sensory memory allows working memory to benefit from these outside systems, as well (Baddley & Larsen, 2007).

Baddeley's model of working memory is particularly relevant to complex tasks such as flying. Flying is a mostly visual-spatial task, and as such, it requires the integration of information that is visually based, and spatial in nature. It also entails a strict task priority hierarchy (i.e., "Aviate, Navigate, Communicate"). The subsystem in the Baddeley's model dedicated to the first order priority tasks involved in flying, such as "Aviate" and " Navigate" is the visuo-spatial sketchpad.

Visuo-Spatial Sketchpad

According to Baddeley's model, the visuo-spatial sketchpad is what is generally known as "visual short-term memory". The visuo-spatial sketchpad is thought to be responsible for encoding and maintaining information relevant to the visual and spatial features of a given set of stimuli. While recognizing that the concept of working memory is one of a multiple component system where no subsystems are functioning in isolation (Baddeley & Larsen, 2007), the emphasis in this dissertation is mostly on the factors and processes involved in air to ground voice communications, that is, the "Communicate" portion of task priority hierarchy of flying, and which specifically involves what is commonly characterized as "verbal short-term memory" - the phonological loop.

Phonological Loop

The phonological loop is assumed to comprise of two components (Baddeley & Hitch, 1994), a phonological store for temporarily maintaining auditory and/or visual verbal input, and an articulatory rehearsal system. In the most recently updated model of the phonological loop (Baddeley, 2003) two pathways are dedicated to the processing of verbal information, one for auditory verbal input (speech), and one for visual verbal input (text). Whereas auditory verbal information is granted automatic access to the phonological store where it enters the rehearsal process, visual verbal information undergoes a different type of processing (e.g., grapheme-to-phoneme conversion and recording) before entering the rehearsal process through the phonological buffer. Traces within the phonological store decay over a short period of time unless refreshed by rehearsal. The rehearsal system is capable of refreshing the memory trace by a general process of attentional activation and reactivation that is available for verbal material as well as for visual, and semantic information. The model assumptions (Baddeley & Larsen, 2007) suggest that the processes of subvocal rehearsal, and the use of subvocalization to name a visual stimulus in order to register it in the phonological store, are in a way less typical than previously thought. Interesting parallel could be drawn between Baddley (2003) and Penney (1989). Penney (1989) argues that within verbal short-term memory, auditory and visual verbal information is processed in two separate streams, which have different properties and capabilities, and the memory trace/code generated by each stream contains different information. Acoustic code is automatic and only generated for auditory inputs. It is sensory-based, and in the absence of subsequent new input, auditory items can be maintained for some time in the verbal short-term memory without conscious attention. Phonological code is generated by the transformation of visually presented sensory input via silent articulation, and the addition to the sensory input of previous knowledge about words, phonemes, and articulation information. Furthermore, according to Penney (1989), the nature of the two types of verbal input also impacts the processing of information these inputs are carrying. Specifically, the acoustic code is hypothesized to be more durable relative to the phonological code, which could explain the very persistent research finding that on short-term memory tasks, auditory presentation almost always results in higher recall than visual presentation, i.e., modality effect. Based on a comprehensive review of the literature, Penney (1989) further argues that auditory verbal information is more robustly organized along the temporal dimension, thus it is remembered better if presented in a sequence, and visual verbal input contents are best associated, and therefore remembered better, when presented at the same time.

In summary, several prominent features characterize the processing of verbal information within working memory:

1. Rapid loss of phonological representations through decay (Baddeley, 2000; Baddeley & Hitch, 1974), where sensory-based inputs generating acoustic code,

in the absence of interference, persist for almost a minute relative to the less robust transformed sensory-based inputs generating phonological code (Penney, 1989);

2. Limited capacity of just a few items unless supported by concurrent rehearsal, and/or by other components of working memory, such as the episodic buffer (discussed elsewhere in this paper) (Baddeley & Larsen, 2007);
3. Exceptional flexibility allowing information presented in any modality to be recorded in almost any other format (Baddeley & Larsen, 2007; Penney, 1989),
4. Better organization of auditory verbal input along the temporal dimension leading to better retention when presented sequentially, and better association of visual verbal input along the spatial dimension and therefore remembered better when presented concurrently (Penney, 1989).

While the original Baddeley's model has been successful in predicting how human cognitive structures function during performance of relatively simple cognitive tasks, the model has encountered some problems especially predicting cognitive functions where more complex cognitive phenomena, not captured by the original model, are involved. For example, the apparent resistance to articulatory suppression specifically in serial recall (Baddeley, 2000), and recall of prose suggested the need to assume a third storage component in the original Baddeley and Hitch (1974) model of working memory - the episodic buffer.

Episodic Buffer

The episodic buffer is a limited capacity system, which provides temporary storage of information in a multimodal code. The buffer is assumed to store episodes with information that is temporarily and spatially integrated (Baddeley, 2002). In that respect it carries many similarities to the notion of episodic memory (Tulving, 1986; 1993). The role of the episodic buffer in Baddeley's model is to integrate information from the phonological loop and the visuo-spatial sketchpad, as well as to serve as the interface between all working memory subsystems and the long-term memory (Baddeley, 2000). Under the control of the central executive, the episodic buffer allows for active maintenance and manipulation of multimodal information such that the integration of working memory and long-term memory is possible (Baddeley, 1996; 2003).

Central Executive

The process of allocating attention and processing resources (e.g., Baddeley, 1996) and the ongoing update of information in the working memory are believed to be essential for the proper interaction between all subsystems in Baddeley's model. Yet, the role of the working memory subsystem responsible for the control of these functions, the central executive, is the least researched and understood component of the model. In essence, the central executive is argued to be an attentional control system accountable for the coordination of ongoing processing tasks (Collette et al., 1999). Its role is critical to the performance of the task at hand as it is responsible for managing the available attentional capacity by focusing, dividing, and switching attention as needed (Baddeley, 2003). In addition, a more recent update of the model (Baddeley & Larsen, 2007) suggests that, as opposed to mostly subvocal, rehearsal within the phonological loop is also an attention-based process, available for different presentation modes, therefore adding a new level of granularity of our knowledge of this important working memory subsystem. In addition, as the focus of research studies moved away from strictly controlled laboratory-based stimuli (e.g., word lists) to real-world stimuli and how people process complex, integrated information, the idea of a fixed executive has become less robust. Real-world environments involve cognitive tasks that require extensive use of knowledge structures (e.g., schemata) from long-term memory. A central executive as described by the original Baddeley's model, although helpful in studying basic cognitive processes in simple cognitive activities, could not provide thorough executive functions in complex knowledge-rich cognitive situations (Sweller, 2005). According to Merrienboer & Sweller (2005), during complex cognitive activities, schemata from long-term memory can act as a central executive by organizing information, or knowledge, that needs to be processed in working memory. This, according to the authors, could promote conditions where working memory, similarly to long-term memory, could become unlimited as well.

Verbal Input Factors

The notions about the structure of the phonological loop, its role as a subsystem in the Baddeley's model, and the differential impact of visual and auditory verbal input characteristics have on processing within the loop, are supported by several groups of research evidence. The findings presented next are relevant to this research because they help better understand the factors influencing the processing within the phonological loop, as well as, identify which of the characteristics of auditory and visual verbal input have the potential to capitalize on the capabilities and minimize the effects of the limitations of the phonological loop.

Modality of verbal input

In the context of short-memory tasks performance, a very robust research finding reported in the literature, and presumed to reflect the inherent structure of working memory, is the so-called “modality effect”. That is, auditory presentation almost always results in higher recall than does visual presentation. For decades now, the modality effect has also been researched (Low & Sweller, 2005; Mousavi, Low, & Sweller, 1995) in the context of multimedia learning environments, and is it discussed later in this dissertation as one of the impact factors in multimedia presentation design.

Per Penney (1989), auditory verbal information is granted automatic access to the phonological store, and the generated acoustic code is more durable than the phonological code generated by visual text presentation which is especially relevant to this dissertation research because it supports the proposition for preserving the role of voice as a primary means for controller-pilot communications message delivery.

Size of verbal input

The “word length effect” is a phenomenon pertinent to the capacity to recall short words better than long words. The longer the word, the longer it takes to say it subvocally. It takes longer to rehearse words with multiple syllables, and to produce them during recall, which allows the memory trace to deteriorate faster (Baddeley, 1966). While not central to this dissertation, the size of verbal input is still a relevant factor that needs to be taken into account when designing any type of verbal display. Specifically, maintaining long ATC messages in working memory could generate excessive cognitive load. In that case, a redundant verbal display may provide a means of preventing such excessive load from occurring.

Interference factors

Irrelevant speech

Another empirically robust working memory research finding is that serial recall performance of visually presented items is reduced by irrelevant speech (Larsen, Baddeley & Andrade, 2000). The irrelevant speech effect is observed regardless of the presentation modality of the items to be recalled and it is equivalent whether the irrelevant speech occurs during or after the presentation of the items. The effect is independent of (a) the phonological similarity and (b) the semantic similarity within the items to be recalled and the irrelevant items (Neath, 2000). Recall is disrupted regardless of the origin (linguistic or not) of the irrelevant material, which suggests that the recall process is operating at the level of speech sound rather than meaning. The phonological loop model explains this effect with the assumption that the irrelevant spoken material is granted direct access to the phonological store, even if participants try to ignore it. Consequently, performance is disrupted as a result of the corrupted memory trace (Baddeley & Larsen, 1994). These findings are relevant to this dissertation research because they support the notion that verbal display redundancy has the potential to be very beneficial in assuring an accurate information transfer in noisy environments such as modern flight decks.

Articulatory suppression

A valuable insight into the processes involved in the phonological loop comes from the finding that if rehearsal is prevented by articulatory suppression the outcome is a significantly degraded performance. When participants are suppressing articulation by being required to repeatedly say an irrelevant sound, they appear to be unable to transfer visually presented material to the phonological store. According to Neath (2000), articulatory suppression eliminates the irrelevant speech effect for visual items.

A number of studies discussed by Baddeley (2000) indicate that the following assumptions can be made about the way information is processed in the phonological loop under the presence or absence of articulatory suppression. First, in the presence of articulatory suppression, incoming auditory stimuli are held in the phonological loop for a few seconds, and then phonologically coded in the multimodal episodic buffer. Visual stimuli are processed in a way very similar to the no articulatory suppression conditions described above, except for the absence of phonological recoding and subsequent articulatory rehearsal. Although performance may be significantly degraded under articulatory suppression, information can still be stored with the help of the episodic buffer (Baddeley & Larsen, 2007). Second, in the absence of concurrent articulatory suppression, auditorily presented items are stored in the phonological loop and maintained using the articulatory rehearsal system, whereas for visually presented items, an additional processing stage is involved in order to allow registration in the phonological store, that is, visual stimuli undergo a grapheme-to-phoneme conversion process and recording. This is accompanied by registration in a separate, multidimensional store, the episodic buffer that is capable of taking advantage, not only of visual and phonological codes, but of syntax and semantics, as well. The factors presented above provide evidence not only in support of the notion that two separate types of processing exist for auditory and visual verbal information within the phonological loop, but also about the extremely adaptive and flexible nature of human memory as a whole. Table 1 summarizes the main factors thought to influence the processing of verbal information according to the Baddley's model of working memory.

Table 1 Factors according to Baddeley's working memory model (Phonological Loop)

Factors	Possible Manipulation	Impact on Processing
Modality of verbal input	Auditory verbal input only, visual verbal input only, auditory + visual verbal input	Auditory verbal input granted automatic access and stronger along the temporal dimension Auditory presentation almost always results in higher recall than does visual presentation Visual verbal input undergoes different processing and stronger along spatial dimension Verbal input presented in both visual and auditory modality may impact processing differentially depending on the temporal relationship and the content (i.e., redundant or not) of the two verbal inputs Differential impact on processing due to interaction with episodic buffer
Interference	Presence or absence of: articulatory suppression; irrelevant speech;	The presence or absence of articulatory suppression affect the way auditory and visual verbal information is processed (e.g., interaction with episodic buffer) Irrelevant speech negatively impacts visually presented items
Phonological make-up of auditory input	Presence or absence of: phonologically similar vs. dissimilar items;	Phonologically similar items are more difficult to remember
Size of verbal input	Short vs. long units of verbal input	Longer units take longer to process, longer to rehearse, faster memory trace deterioration

Cognitive Load and Multimedia Learning

The cognitive load theory (Sweller, 1988, 1994, Chandler & Sweller, 1991; Sweller, Van Merriënboer & Paas, 1998) and the cognitive theory of multimedia learning (Mayer, 1997, 2001; 2005a; 2005b) utilize knowledge about human cognitive architecture, and argue that information should be presented such that the limited working memory capacity (Baddeley, 1992; Chandler & Sweller, 1991) is used as efficiently as possible. While the cognitive load theory is primarily focused on the load imposed on working memory during instruction, and the factors that influence the conscious information processing during the performance of a specific cognitive task, the theory of multimedia learning is mostly concerned with the potential risk of cognitive load associated particularly with multimedia learning environments. The central themes of these two models are relevant to this dissertation because they transform the basic concepts from Baddeley's working memory model into a set of practical guidelines, which then allow a direct application of these guidelines into the design of information displays.

Cognitive Load Factors

Schemata

While studying the perceptual structures that chess players perceive after successive glances at the chess position, Simon & Chase (1973) used the term "chunk", also identified as schema, to characterize how expert chess players develop, and use their exceptional memory. In the cognitive load theory, schemata are a type of organized knowledge structures in long-term memory that represent objects, situations, and events, and allow the categorization, understanding, and use of incoming information appropriately (Sweller, 2005). Schemata acquisition and automation are the primary mechanisms of learning. Schemata are initially associated with specific situations from which they originate. With experience, they are increasingly associated with general principles, and become organized into large knowledge structures. Schemata vary in their degree of complexity and level of automation, and operate under controlled (when the information needs to be consciously attended), or automatic processing (occurs automatically without conscious effort) (Chase & Simon, 1973). Ultimately, the goal of learning is to store automated schemas in long-term memory and therefore allow rapid individualized access to them as a critical component of any skilled performance.

Unlike in the virtually limitless in capacity long-term memory, where elements are stored in a form of hierarchically organized schemata (thus allowing the processing of large amounts of information), the limited capacity of working memory is a challenge. However, the process of schemata acquisition and automation can effectively modify the characteristics of working memory. According to van Merriënboer and Sweller (2005), handling even a very complex schema as just one element can significantly reduce the cognitive load imposed during instruction. Also, automated processing requires less working memory processing capacity. Moreover, by organizing the information that need to be processed in working memory, schemata can effectively play the role of a central executive, and promote conditions where working memory can be practically unlimited (Sweller, 2005).

Types of Cognitive Load

Cognitive load theory purports that since working memory is limited, and, if the complexity of instructional materials is not properly managed, this may result in a cognitive overload which can hinder schemata acquisition, and later result in a degraded performance

(Sweller, 1988). The theory identifies three types of cognitive load: (a) intrinsic, (b) extraneous, and (c) germane (Sweller, 1988; Sweller, Van & Paas, 1998; Sweller, 2010). According to Sweller and his colleagues, intrinsic cognitive load is "the mental work imposed by the complexity of the content" (Clark, Nguyen, & Swelller, 2006, p. 9). This type of cognitive load is inflicted by the inherent complexity of the information rather than by instructional design, and represents the essential amount of processing resources that are required to understand the material (Chandler & Sweller, 1991). The two types of cognitive load particularly associated with the presentation format of instructional materials, are extraneous cognitive load (Chandler & Sweller, 1991; Chandler & Sweller, 1992), and germane cognitive load (Sweller, van Merriënboer, & Paas, 1998). Because intrinsic and extraneous load are additive, extraneous cognitive load can reduce instructional effectiveness only when coupled with a high intrinsic cognitive load. If total cognitive load is not excessive due to a relatively low intrinsic cognitive load, a high extraneous cognitive load may not be a concern because learners can easily assimilate low element interactivity materials (Paas, Renkl, & Sweller, 2003; Paas, Tuovinen, Tabbers, and van Gerven, 2003). The third type of cognitive load is germane load. This type of cognitive load is related to the remaining free capacity in working memory that can be redirected from extraneous load toward schema acquisition (Sweller, van Merriënboer, & Paas, 1998).

Complexity

More recently, the cognitive load theory was updated by a series of publications (Paas, Renkl & Sweller, 2003; Sweller, 2010) introducing the notion that intrinsic cognitive load depends on the level of element interactivity or complexity of instructional materials. That is, information content of instructional materials varies on a continuum from low to high in element interactivity, and the different levels of interactivity cannot be altered by design. The level of element interactivity refers to the extent to which individual information elements can be understood and learned without having to learn the relationship between any other elements (van Merriënboer & Sweller, 2005). Element interactivity is low when each individual element can be understood, and learned without referencing other elements, and high when material cannot be fully understood until all of the elements, and their interactions (especially when elements are syntactically and semantically connected) are processed simultaneously (Paas, Renkl & Sweller, 2003).

Most cognitive load theorists consider complex, high element interactivity materials, which require the understanding and learning of multiple elements and their interconnections, as schemata; and low element interactivity, simple to learn materials as individual elements rather than schemata. Furthermore, they consider the failure to assimilate all the elements of high element interactivity material as equivalent to a failure to understand the concept as a whole, while failure to assimilate all the elements of low element interactivity material, as equivalent to a failure to learn or remember. Therefore, understanding can be defined as the learning of high element interactivity material. In other words, low interactivity material needs to be merely learned (and not understood, and learned) (Sweller, 1994). The concept of intrinsic cognitive load is relevant to not only learning but also to any complex cognitive task that involves real time processing of new information that contains elements with different levels of interactivity. For example, in the context of this dissertation research, ATC instructions may vary in the number of information elements (from a single altitude assignment to multiple-element assignment that includes elements such as altitude, speed, heading, radial, etc.), as well as, in the level of interactivity between these elements. More specifically, ATC procedures and airspace rules are designed to maintain aircraft separation (the distance by which aircraft avoid obstacles or other aircraft). ATC services issue flight clearances based on route, time, distance, speed, and altitude. An ATC clearance is a highly complex navigation solution accounting for all these factors and communicated to the pilot in a very structured procedural instruction. A clearance always specifies a clearance limit (the farthest the aircraft can fly without a new clearance) and is typically followed by a heading or route to follow, altitude, communication frequencies, and transponder codes. ATC may also assign headings, also known as radar vectors. Radar vectors are another method used by ATC to provide separation between aircraft for landing, especially in busy traffic environments. Therefore an ATC clearance (especially an IFR clearance) is a complex instruction with highly interactive elements that are syntactically and semantically connected. For example, "*Gulfstream 7552, cleared to Stockton Airport via turn right heading zero-six-zero within one mile of the airport. Radar Vectors San Jose, then as filed. Maintain 3,000 expect 5,000 five minutes after departure. Departure frequency is 121.3, squawk 426*", is a an IFR clearance that contains multiple information elements including:

1. Specific call sign.
2. Clearance limit: the farthest destination the aircraft is allowed to go under IFR.
3. That the pilot is expected to execute the right turn to 060°, without further ATC prompting, within one mile of the departure airport.
4. The departure controller will provide directional guidance to the San Jose VOR.

5. After arriving at the San Jose VOR, the pilot will likely resume navigation without ATC prompts along the airways and intersections that were filed in their flight plan.
6. After takeoff, the pilot is expected to climb to an altitude of 3000 feet above sea level.
7. The final altitude assignment is probably going to be 5000 feet above sea level. However, the pilot must follow actual ATC altitude assignments throughout the flight. This portion of the clearance provides a backup if communications are lost, allowing pilot to proceed to climb and maintain 5000 feet.
8. After Palo Alto, Tower instructs the pilot to contact "Departure" on the specified communication frequency.
9. The aircraft transponder should be programmed to 4263 so that ATC can positively identify it on radar.

In summary, the elements of an ATC IFR clearance are highly interactive and in order to safely complete the task, all of the elements and their interactions must be understood simultaneously as a whole.

Format

In contrast to intrinsic cognitive load, extraneous cognitive load results from any redundant or superfluous features of the instructional material, and is generated by the format in which information is presented to learners (Pollock, Chandler & Sweller, 2002). While the intrinsic load is thought to be un-modifiable, designers can modify the format of instructional material in order to reduce extraneous cognitive load. Eliminating such features allows both keeping the extraneous load as low as possible, and directing all available working memory resources to learning (Sweller, van Merriënboer, & Paas, 1998). Since extraneous cognitive load and intrinsic cognitive load are additive, if intrinsic cognitive load is high, extraneous cognitive load should be reduced, and if intrinsic load is low, a high extraneous cognitive may not impede learning because the total cognitive load is within working memory limits. However, traditional methods of reducing extraneous cognitive load, especially when applied to complex learning tasks have not been very successful in lowering the total cognitive load to an acceptable level (van Merrienboer & Sweller, 2005). Quite often very complex learning tasks generate excessive cognitive load that may seriously impede learning especially for novice learners, therefore not leaving enough cognitive resources for schemata construction and automation. As a result, more recently, rather than only trying to reduce extraneous load, researchers began to explore new instructional methods that affect intrinsic and modify extraneous cognitive load (Merrill, 2002). These new methods are focused on the transformation of extraneous load into germane load which helps the construction of schemata by redirecting the attention of the learners to only directly relevant material (Sweller, van Merriënboer, & Paas, 1998).

The manner in which information is presented to learners, and the required learning activities are all factors relevant to levels of germane cognitive load. Because germane load can actively support schemata acquisition and automation, this transformation may effectively reduce the intrinsic cognitive load. Ultimately, when intrinsic cognitive load is reduced, the total cognitive load is reduced, therefore freeing working memory capacity. The freed working memory capacity permits the use the newly learned material for the acquisition of more advanced schemata, knowledge, and skills (Sweller, 2010).

In essence, extraneous cognitive load can clearly be identified as the load associated with all the non-essential features added to the presentation of information without enhancing its central function. The transformation of such features into germane cognitive load could be critical, in my view, for the success of any design aimed at optimizing performance of complex cognitive tasks involving processing of new information. More specifically, this dissertation research will explore one method of transforming extraneous into germane cognitive load by modifying the presentation format such that it would enhance rather than impede the processing of new information within working memory. The format modification includes a temporal offset, redundant mode, and a presentation layout (for the redundant mode). The intent of a temporal offset (i.e., non-concurrent) is to reduce the potential for high extraneous cognitive load generated by a concurrent presentation of voice and text, while the redundant mode (text) is aimed to mitigating the effects of poor intelligibility associated of the primary delivery mode (voice). The use of an automatically generated, well established shorthand layout for the text presentation targets the lack of existing or fully developed schemata for less experienced users and helps automate already existing schemata for more experienced users.

Complexity X Format

The total cognitive load is a mixture of two factors, (a) complexity, or intrinsic cognitive load, which is determined by element interactivity and (b) format, or extraneous cognitive load, which is artificially imposed by the method of information presentation. Therefore, because complexity and format have cognitive load consequences, the relationship between these factors needs to be considered. The notion that intrinsic cognitive load cannot be modified has important implications for instructional design and information presentation in general. The combined consequences of a high extraneous and high intrinsic cognitive load may overwhelm the limited processing capacity of working memory. Inappropriate design can impose a heavy extraneous cognitive load and if intrinsic cognitive load is already very high due to high element interactivity, the total load can exceed cognitive resources, leading to a learning failure. The interaction of complexity and format is very relevant in the context of this dissertation research. Historically, the high element interactivity in an ATC IFR clearance has been handled by training or procedures. Specific format requiring a certain order of delivery and consistent clearance structure has been mandated. Pilots shorthand clearances using a specific shorthand technique they learn from their instructors. This approach however has had a limited success in reducing controller-pilot communication errors due in part to high clearance complexity (especially of some IFR clearances) coupled with a delivery format that is not always consistent, and along with the requirement for the pilot to shorthand the clearance in order to have a somewhat permanent record of it. An additional source of extraneous cognitive load, atypical for a learning environment but all too familiar for controllers and pilots engaged in voice radio communications, is the poor intelligibility associated with the incoming ATC messages. This dissertation research will investigate how the high extraneous cognitive load imposed by voice radio communications may be transformed into germane load by employing a redundant format for information presentation.

Table 2 summarizes the main factors thought to influence the processing of information according to the cognitive load theory.

Table 2 Factors according to the cognitive load theory

Factors		Possible Manipulation	Impact on Processing
Cognitive load	Intrinsic (Complexity)	Different levels of inherent complexity and interactivity of information elements	When element interactivity is high, material cannot be fully understood until all of the elements, and their interactions are processed simultaneously
	Extraneous (Format)	Information presentation containing different types of superfluous features	High extraneous load has negative impact on processing
	Germane	Presence or absence of features supporting transformation of extraneous into germane load	Can actively support schemata acquisition and automation, ultimately can reduce the intrinsic cognitive load
Schemata		Identifying the presence or absence of previous knowledge	Can play the role of a central executive, and promote conditions of practically unlimited working memory

Multimedia Factors

While the cognitive load theory is mostly focused on managing cognitive load, and facilitating the building and automation of schemata, the cognitive theory of multimedia learning adds a level of granularity to that by taking into consideration the factors associated with the cognitive load generated by multimedia environments. More specifically, in such environments the building of mental representations by integrating verbal and visual information in working memory is a critical step in the learning process. Furthermore, the switching between visual and verbal instructions to mentally integrate them is a very cognitively demanding task accomplished at the expense of mental resources that could otherwise be allocated to the learning process (Mayer, 2001). The potential risk for cognitive overload in a multimedia-learning environment, and the means to reduce that risk are central to the cognitive theory of multimedia learning (Mayer, 1997; Moreno & Mayer, 2002). The notion that modern flight decks are indeed environments very rich in multimedia, and the knowledge about the factors discussed in the next section of this dissertation were fundamental in determining the design features of the redundant verbal display proposed for investigation in this research. Also, because these factors are so closely intertwined, and difficult to understand in isolation, their relevance to this dissertation becomes apparent only when considered as a group. Interestingly, for over 20 years now, the effects of several factors have been consistently demonstrated by research conducted in both the cognitive load theory and the cognitive theory of multimedia learning domains.

Factors Common in Cognitive Load and Multimedia Learning Domains

Spatial and temporal contiguity

When sources of information are separated in space or time, and are also difficult or impossible to understand in isolation, in order to understand the material, learners must to split their attention between these sources (i.e., the split-attention effect) (Mousavi, Low, & Sweller, 1995; Tindall-Ford, Chandler, & Sweller, 1997). For example, learners must hold segments of text in working memory while searching for the matching visual representation until information becomes comprehensible. Such process inhibits learning because it involves mental integration of the material, and requires working memory resources that would otherwise be available for acquisition of schemata. However, when the multiple sources of information are physically integrated (Chandler and Sweller, 1991, 1996; Mayer, 1989; Mayer & Gallini, 1990) or close to each other rather than spatially separated (i.e., the spatial contiguity principle) (Moreno & Mayer, 1998; 1999; Mayer, 2001), understanding can occur without an unnecessary visual search. Consequently, cognitive load may be reduced, and the acquisition of schemata facilitated.

Dual-mode presentations (Paivio, 1991) have been recommended as a potential alternative to spatial contiguity for mitigating cognitive load issues associated with split-attention, that is, it is better to present an information material using two modes rather than one (Mayer & Anderson, 1991, 1992). However, Mayer and Sims (1994) suggested that dual-mode instructions may be superior only when the audio and visual information are presented simultaneously rather than sequentially (i.e., temporal contiguity principle) (Mayer & Anderson, 1991; 1992). Specifically, learners understand the material better when corresponding words and visuals are presented at the same time than when they are separated in time. In the cognitive load theory, the temporal contiguity effect is described as a special case of split attention (Mousavi, Low & Sweller, 1995) however both theories agree that in split-attention conditions, physically and temporally integrating separate sources of information, or using more than one modality, all produce a positive effect on learning (Jeung, Chandler & Sweller 1997).

Redundancy

The use of fewer rather than many extraneous words and visuals when employing multimedia for presenting information is a founding principle of the cognitive theory of multimedia learning (i.e., the coherence/redundancy principle) (Mayer, Heiser & Lonn, 2001). When different sources of information are intelligible in isolation, and each source provides identical content only in a different form, an unnecessary cognitive load is imposed by the mere existence of multiple redundant sources of information. The cognitive load theory refers to this as the redundancy effect (Sweller, van Merriënboer J. & Paas, 1998). Attending to any superfluous / redundant information in multimedia instructions increases the extraneous cognitive load, because part of the working memory capacity is used for the processing of unnecessary information that does not contribute to learning and construction of schemata (Chandler & Sweller, 1991; Sweller, 1994).

Cognitive load theorists identify several types of redundant information. First, if particular information can be derived from other elements in the instructional material, concurrently presenting the same information in multiple forms can have a neutral, or even negative, effect on learning (Kalyuga, Chandler & Sweller, 1999; Mayer, Heiser & Lonn, 2001; Mousavi, Low & Sweller, 1995). Attending, and processing narration and text concurrently, along with relating them to visual materials, requires additional cognitive resources, overloads working memory, and ultimately hinders learning. Without the presence of visual depictions, a concurrent presentation of identical auditory and visual text is significantly less efficient in comparison with a narration only presentation. Sequential presentations, however, allow both modes to be handled without working memory overload, with the second presentation being used as reinforcement of the positive effects of the first presentation (Kalyuga, Chandler & Sweller, 2004). Second, information irrelevant to learning, and added only to embellish the multimedia instruction or make it more interesting and engaging, is superfluous and does not contribute to learning (Harp & Mayer, 1998; Mayer, Heiser & Lonn, 2001). Third, over time as learners develop expertise in a particular domain and consequently information they are already familiar with, can become redundant (Kalyuga, Chandler & Sweller, 1998). A source of information that may be essential for a novice could become redundant for someone with more knowledge in a particular domain, and the cognitive load effects can first diminish and then be reversed with additional experience (Kalyuga, Ayres, Chandler & Sweller, 2003).

Modality

In multimedia environments information is frequently presented in different modes. Consequently, the presentation modes impact the way it is processed because of the different modalities involved. The phenomenon associated specifically with the positive effect of employing more than one modality in multimedia instruction is referred to as the modality effect (Sweller, van Merriënboer J. & Paas, 1998). Modality effect (in the cognitive load theory) (Sweller, van Merriënboer J. & Paas, 1998) or modality principle (in the cognitive theory of multimedia learning) (Mayer, 2001) is associated with the finding that learning is more effective when the visual materials are accompanied by narration rather than by text (Jeung, Chandler & Sweller, 1997; Mousavi, Low & Sweller, 1995; Mayer & Moreno 1998). The traditional explanation of modality effect suggested that when both modalities are utilized, the working memory capacity is increased. More recently, Tabbers, Martens, and van Merriënboer (2004) challenged this view. Based on the working memory architecture, they argue that modality effect demonstrated in earlier research cannot be accounted for in terms of an increase in available working memory resources. Specifically, per Baddeley's model, the phonological loop is responsible for processing of verbal information, and the visuo-spatial sketchpad is responsible processing of visual and spatial information (Baddley, 1992; 1998; Baddeley & Larsen, 2007). Therefore, words presented as text, or speech, are processed in the phonological loop in spite of the mode, and only when this articulation process is interrupted (e.g., irrelevant speech, and articulatory suppression), phonological code is not produced. For that reason, replacing visual text with narration in multimedia instructions may not necessarily increase the total working memory capacity (Tabbers, Martens & van Merriënboer, 2004). Rather, the working memory capacity is utilized more efficiently by reducing visual search (when text is presented as a narration) (Jeung, Chandler & Sweller, 1997; Mousavi, Low & Sweller, 1995; Tindall-Ford, Chandler & Sweller, 1997), and better temporal contiguity (Kalyuga, Chandler & Sweller, 1999). In summary, according to Tabbers, Martens and van Merriënboer (2004), the positive effects on learning when two modalities are engaged can be explained not by the superiority of narration over visual text but by the optimal integration of text and visuals that prevents learners from splitting their attention.

When the inherent complexity of the material is high (i.e., high element interactivity), the split-attention, and redundancy effects are readily demonstrated, however they disappear when low element interactivity material is used (Chandler & Sweller, 1996; Sweller & Chandler, 1994). In addition, Tindall-Ford, Chandler, and Sweller (1997) found that the modality effect could only be obtained using high element interactivity material as well. The finding that cognitive load effects can only be obtained using instructional materials with high complexity is defined as the element interactivity effect. It consists of an interaction between the split-attention, redundancy, and modality effects; and the complexity of the material being learned (Sweler, 2005).

Expertise

The interaction between the basic cognitive load effects (e.g., split attention, modality, and redundancy effects) and the level of expertise represents the so-called expertise reversal effect. The effect is demonstrated when instructional methods that work well for novice learners have no, or negative, effect with more experienced learners (Kalyuga, Chandler & Sweller, 1998). Experts store a large number of domain specific schemata, which allows them to organize many elements of related information into a single element. In that respect, the level of expertise is a critical factor in determining what information is relevant when designing instruction. For example, experts are able to recognize a pattern in a set of elements as a known schema, and treat the whole configuration as a single unit. A single, high-level element requires considerably less working memory capacity for processing than the many low-level elements it contains. As a result, the schemata, stored in long-term memory, allow experts to avoid processing large amounts of information and effectively reduce the load imposed on limited capacity working memory. Furthermore, many of the expert schemata are highly automated due to extensive practice, which supports the notion that higher level of expertise in a particular domain is an important means of reducing cognitive load (Van Merriënboer & Sweller, 2005). If information has become redundant due to an increased expertise in a certain domain, and such information is nevertheless provided, the experts may not be able to avoid attending it. As redundant information is often difficult to ignore, it still requires working memory resources for processing, and that may cause a cognitive overload. For that reason, elimination rather than integration of redundant sources of information may be more beneficial for experienced learners. In summary, the most important implication of the expertise reversal effect is that, in order to be efficient, instructional design should be tailored to the learners' level of expertise (Kalyuga, Ayres, Chandler & Sweller, 2003).

Pace

Tabbers, Martens & van Merriënboer (2004) reported that modality effect could not be replicated in their study and suggested that it does not easily generalize to non-laboratory, more ecologically valid environments. They explain the reported reversal of modality effect with the use of learner-paced instructions as opposed to the system-paced instructions used in previous research by Mayer and Moreno (1998) and Kalyuga, Chandler and Sweller (1999), therefore concluding that dual-mode instructions may only be beneficial when multimedia instruction is system-paced, and visual-only instructions are more effective when instruction is self-paced. The advantage of dual-mode in system-paced instruction is that pictures and text can be perceived simultaneously, resulting in a lower extraneous load than in visual-only instructions where the learner has to switch between text and picture in a fixed period of time. In self-paced instructions, however, this advantage disappears because the learner with the visual-only instructions has more time to relate the text to the picture. In addition, with visual text it is much easier to browse through than with narration, which is inherently linear. Consequently, self-paced instructions could make visual-only instructions more effective than dual-mode instructions, and actually reverse the modality effect (Tabbers, Martens & van Merriënboer, 2004).

The cognitive load and multimedia factors discussed above encapsulate the basis for the specific design solution proposed for this dissertation research. A subset of the following guidelines for presenting verbal information in multimedia environments derived from the review of literature was applied:

- When the presentation is controlled by the system (system-paced) and its content is relevant and non-redundant to the visual material it supports; verbal information should be presented as (a) narration rather than text, and (b) concurrently with the pertaining visual material
- When the presentation is self-paced, its content is relevant, and non-redundant to the

visual material it supports, verbal information should be presented as text only

- When the visual material is intelligible by itself, or has become redundant with users' advancing expertise, both narration and text should be removed, or made available at the users' discretion
- When no visual material is present, identical in content narration and text should be presented sequentially, or the one presented second should be made available at the users' discretion.

A summary the main factors thought to influence the processing of information according to the Cognitive Theory of Multimedia Learning is presented in Table 3.

Table 3 Factors According to the Cognitive Theory of Multimedia Learning

Factors			Possible Manipulation	Impact on Processing
Contiguity	Spatial Temporal		Different spatial and temporal organization of information	Lack of contiguity leads to split-attention and high cognitive load; slows down processing;
Modality	*Visual-spatial*		*Visual-spatial* + visual verbal input; *Visual-spatial* + auditory verbal input; *Visual-spatial* + auditory verbal input + visual verbal input	More effective when visual materials are accompanied by narration (i.e., visual-spatial + auditory verbal);
	Verbal	Visual Auditory		
Redundancy	Same information in multiple forms Superfluous and irrelevant Becomes unnecessary as learning progresses (i.e., with higher levels of expertise)		Different types and/or levels of information redundancy	Imposes high cognitive load, slows down processing; With non-concurrent presentation the type of redundancy where the same information is presented in multiple forms can serve as a reinforcement of the positive effects of the information presented first;
Pace			System-paced vs. self-paced	When system-paced dual-mode presentation benefits from modality effect; when self-paced visual only presentation (graphical + text) is more beneficial;
Expertise			Different levels of expertise	Helps diminish, and then reverse all cognitive load effects;

Individual Differences

After many years of instructional design research, scientists have found that even after the most careful application of proper design principles, multimedia learning environments tend to help some learners more than others. For example, the effectiveness of combining pictorial and verbal information may vary depending on instructional content and learners' individual differences such as verbal and spatial ability, prior knowledge, etc. (Kalyuga, Chandler & Sweller, 2000; Moreno & Mayer, 1999; Mayer & Sims, 1994). Yet, the relationship between individual differences and learning from multimedia representations remains understudied (Moreno & Plass, 2006; Mayer, 2001).

The dimensions of individual differences found to have moderating, or even mediating effects on learning and cognitive performance outcomes include prior knowledge (Ackerman & Beier, 2005), spatial and verbal ability (Plass, Chun, Mayer & Leutner, 2003; Moreno & Plass, 2006), learning preferences (e.g., visualizer vs. verbalizer) (Leutner & Plass, 1998), cognitive styles and strategies, and affective factors (Graff, 2005; Sadler-Smith & Smith, 2004). While most of these factors do have an effect on learning and performance in general, for this dissertation research, a review of those particularly associated with cognitive performance in multimedia environments (i.e., prior knowledge, verbal and spatial ability, and learning preferences) is presented next.

Prior Knowledge / Level of Expertise

In complex, multiple-task environments, the effectiveness of learning and cognitive performance in general, is influenced by the processing limitations of working memory. Prior domain-specific knowledge and the associated levels of expertise are considered a primary means for reducing these limitations as well as managing complex cognitive activities (Kalyuga, Chandler & Sweller, 2000). Therefore, understanding the role of prior domain knowledge is critical for the successful management of cognitive load in such environments.

Most cognitive activities occur in specific domains, and are based on, and managed by, domain-specific schemata. These schemata allow quick encoding, and storage, of large amounts of information in long-term memory (Sweller, 2005; 2010; van Merriënboer & Sweller, 2005). Because of the major differences in the amount, and levels of schemata automation, novices and experts process the same information very differently. A number of studies have found that domain expertise is defined by the larger, and better, set of schemata that experts possess (Kalyuga, Chandler & Sweller, 1998). They rely on the retrieval and activation of these schemata when performing tasks within their area of expertise. For them, there are no severe working memory limitations for knowledge-based performance. In contrast, in the absence of relevant prior knowledge, novices have to process many new elements of information that may lead to increased cognitive load. While for the experts all necessary knowledge structures are available in long-term memory, external guidance (e.g., guided instruction) may be the only available source of executive function for novices (Kalyuga, 2005).

Accordingly, the design of instructional materials, or information presentation in general, intended to support the performance of a particular cognitive task, should account for the already existing schemata (i.e., level of expertise), and balance it with direct external guidance (e.g., additional instruction or sources of information). In other words, an executive function should be based on the existence of knowledge necessary for dealing with familiar and previously learned components of incoming information, and on instructional guidance only when required for dealing with new, unfamiliar information (Kalyuga, Ayres, Chandler & Sweller, 2003).

In summary, expertise is characterized by the large amount of schemata that experts can access. The availability of these knowledge structures can effectively reduce the processing limitations of human cognitive system, and fundamentally change the characteristics of performance. These structures direct the allocation of cognitive resources and significantly influence the processing of multimedia materials.

Verbal and Spatial Ability

For the optimal design of multimedia learning environments, which require the processing of verbally, visually, and spatially encoded material, it is important to account for the learner's verbal and spatial ability. Studies have found verbal ability to be a predictor for the effectiveness of visual aids. For example, Levie and Lentz (1982) concluded that low-verbal ability learners might benefit from visual aids more than for those with high-verbal ability. Moreno and Plass (2006) reported that verbal ability was the only predictive measure of learning outcomes, and indicated that because of the strong association found between verbal ability and intelligence, general ability maybe the sole factor that can help explain individual differences in multimedia learning environments.

Several studies of "Attribute X Treatment Interactions"(ATI) showed that for low-prior knowledge and high-spatial ability students, multimedia effects are strongest. For example, Mayer and Sims (1994) found that high-spatial ability students are more likely than low-spatial ability students to build mental connections between visually based and verbally based representations. They concluded that pictures synchronized with words would be most beneficial for high-spatial ability students. Plass, Chun, Mayer, and Leutner (2003) investigated the role cognitive load plays in multimedia learning environment, and more specifically, how cognitive load interacts with learners' cognitive abilities when processing verbal and visual information in such environments. The results suggested that learners should have options for using study material in both visual and verbal mode but should not be presented with all available information, and forced to process it, all at the same time.

In summary, providing options for visual and verbal modes is only effective in addressing individual differences when learners can choose which information they would like to select and process. This could be implemented in practice by providing features that allow requesting information instead of presenting it by default to all users (Leutner & Plass, 1998). A summary of the individual differences factors thought to influence the processing of information in multimedia environments is show in Table 4.

Table 4 Individual differences factors

Individual Differences Factors	Possible Manipulation	Impact on Processing
Prior Knowledge/Level Expertise	Novices, intermediate, and experts	The existence of large amounts of knowledge structures reduces the demand on WM when processing new information;
Verbal Ability	Post-hoc grouping of high vs. low verbal ability users	High-verbal ability helps holding and manipulating verbal information in working memory, as well as building mental representations based on text alone;
Spatial Ability	Post-hoc grouping of high vs. low spatial ability users	High-spatial ability helps creating, holding, and manipulating spatial representations in working memory as well as elaborating images that express the content of text; High-spatial ability helps better recall of text that evokes imagery;

Synthesis

Two main factors influence the ease with which instructions or procedures are understood: (a) the intrinsic complexity of the information and (b) the way information is presented. Understanding and carrying out instructions or procedures is a complex cognitive activity where these factors interact with (a) each other, (b) the relevant characteristics of human information processing system, and (c) the environment where such activity is performed. Balancing these interactions is essential for the utility and usability of any system designed to support human performance. For example, the intrinsic nature of some information may not allow handling it in a serial fashion because of the high element interactivity. If schemata exist in long-term memory, the cognitive load imposed on working memory will be low and understanding high, and vise versa under conditions where no schemata exist. Similarly, if the presentation format adds an unnecessarily high number of elements that need to be processed simultaneously, this can dramatically increase extraneous cognitive load and hinder understanding, especially when the intrinsic cognitive load is already high. However, if the presentation format is utilized as a feature promoting the creation and automation of schemata by reducing the number of elements that need to be processed simultaneously, then even already intrinsically high-complexity material can be handled with ease.

Today's advanced technology affords virtually unlimited amount of information to be presented to a user in a multimedia fashion. This is particularly evident in modern flight decks and more that ever, especially when it comes to visual information. Yet, to this day, the voice radio communication system relies mostly on the pilots' training and experience, as well as operational procedures, as the balancing act against the potential for miscommunication between controllers and pilots. For the period of time between receiving and acting upon the controller's instructions, pilots must shorthand, thoroughly understand, and retain the information in his/her memory. Furthermore, pilots' shorthand notes are the only permanent record of the controllers' message on the flight deck. The added effects of ATC message complexity, the multiple ongoing flight-related tasks, especially in traditionally high workload phases of flight (e.g., approach and landing), the format in which these messages are presented, and the typically low intelligibility of voice radio communications which is not confined to only older flight deck designs, have the potential to impose excessive pilot cognitive load.

Here, I propose adding a layer of redundancy to the display of verbal information which can help limit pilot's cognitive load to only the essential amount of processing resources (i.e., intrinsic cognitive load) required to understand the information conveyed by the controller as well as eliminate the need for the pilot to split his/her their attention between listening and trying to capture the information by writing it down while performing other ongoing tasks such as monitoring navigation or system performance information. This solution involves an automatic, non-concurrent voice-to-text (V-T-T) transcription of the controller's message, which would be available for a review on a visual display, immediately after it is delivered by voice, or at any time during a flight. The decision to propose a non-concurrent (as opposed to a concurrent) display of the same verbal information content in two different modes is based on the empirical evidence discussed in the literature review and the design guidelines derived from it. In essence, the goal is to provide an alternative means for capturing and retaining the information conveyed by voice without generating extraneous cognitive load and by minimizing or eliminating the potential for redundancy effect.

More specifically, drawing upon the most recently updated model of the phonological loop (Baddeley & Larsen, 2007), at the onset of the auditory verbal input (voice), the pathway within the Phonological loop, which responsible for processing of auditory verbal information is engaged, and auditory information is granted its direct access to the phonological store. At that time, all attentional resources are focused on actively processing the auditory input. Then, following the auditory information delivery, the visual verbal input is presented (V-T-T transcription) and the pathway responsible for the processing of visual verbal information is activated. The visual verbal input enters the rehearsal process through the phonological buffer after the additional processing required for it is complete. At that point, all attentional resources have switched to processing of visual verbal input. This very short temporal offset helps reinforcing the memory trace "left" by the auditory verbal input and may be especially useful when the auditory input is very complex or distorted (as it frequently is with voice radio communications between controllers and pilots). In addition, the availability of a permanent record for a review at a later time, as well as, merely the awareness of the existence of such record can be beneficial from stand point of flight operations and management of pilot's cognitive resources.

Based on these most recent updates of the cognitive load theory (Paas, Renkl, & Sweller, 2003; Sweller, van Merrienboer, & Paas, 1998; Van Merriënboer & Sweller, 2005) and the cognitive theory of multimedia learning (Mayer, 1997, 2001; 2005a; 2005b), multimedia environments should only incorporate information that contributes to the creation and automation of schemata, and omit all redundant information. Furthermore, the format in which the information is presented should facilitate the processing with working memory for novices who are less experienced and help automate already existing schemata for more experienced users. In my view, the application of these guidelines to a non-learning environment (e.g., an operational environment) could be as beneficial in optimizing the performance of complex cognitive tasks but only after a careful examination of the inherent attributes of such environment. More specifically, an operational environment may be very different when it comes to: (a) the quality of the sensory input and the visual and acoustic ambiance as a whole, (b) the ability of the user to control the pace and timing of information input, (c) the need for the user to attend other ongoing tasks, or (d) how time-critical or safety-critical the task is. The differences between a learning environment and an operational environment such as a flight deck based on these attributes are highlighted in Table 5 below.

Table 5 Differences between a learning environment and a flight deck

Attributes	Learning Environment	Flight Deck Environment
Quality of sensory input	• High quality audio-visual presentations; • Free of noise and interference; • Controllable;	• The incoming controller-pilot communications audio is often with very poor quality; • Very limited control available;
Quality of visual and acoustic environment	• Noise and interference- free; • Controllable;	• Often very noisy acoustic environment; • Somewhat controllable;
Task environment	• Single-task;	• Multiple-task;
Pace, timing, complexity of input	• Predictable and very controllable;	• Not very predictable with limited control of pace and timing;
Task criticality (time or safety)	• Low;	• High;

Although this list of attributes is not exhaustive, it can nevertheless help the deployment of learning environments' design guidelines in the design of operational enviroments of without violating their underlining principles. For example, the guideline prescribing the omission of all redundant information for a learning environment may be modified to “allow” certain level of redundancy be added to the display of verbal information on a flight deck. Specifically, the implementation of an automatic V-T-T transcription of the voice communication content may help minimize the effects of high complexity of the verbal input and poor overall acoustic quality of the ambiance. Furthermore, a sequential presentation of the voice message and the transcription may eliminate the potential for redundancy effect, which is predicted to occur if they were to be presented simultaneously (Kalyuga, Chandler & Sweller, 2004). In addition, in support of the high criticality associated with voice communications, the V-T-T transcription could also be used as: (a) a means to verify the content of a clearance, and (b) a memory aid available for access at any time during a flight.

When considered in the larger context of flight deck operations, the information received via controller-pilot communications is always in support of what is typically visual-spatial information (e.g. a navigation map, a chart, etc.), associated with the task of piloting the aircraft (i.e., “Aviate” and “Navigate”). In that respect, the guidelines discussed above are also relevant, and when applied, may help optimize the integration of the verbal and visual-spatial information, so that both, the potential for a higher cognitive load and the conditions conducive to communication errors, are minimized, and ultimately eliminated by design.

Factor Categorization

For the purpose of this dissertation, Table 6 below identifies the various factors into variables that are manipulated, randomized, covaried-out, or fixed. A specific description of each selected manipulation along with the respective hypothesized effect on the utility of voice-to-text transcription of the controller-pilot voice communications in managing pilot cognitive load are described in the next section.

Table 6 Proposed factor categorization

Factor			Experimental design	Hypothesis/Justification
Cognitive load	Intrinsic		*Suggested manipulation:* Low, medium and high intrinsic complexity of information content;	The higher the intrinsic complexity of information the higher the intrinsic cognitive load;
	Extraneous		*Fixed at:* *V-T-T transcription formatting with use of shorthand abbreviations;*	Shorthand abbreviations provide structure and organization to the transcript which help keeping the levels of extraneous cognitive load low;
	Germane		*Randomize*	Germane load is not assessed in this dissertation;
Size of verbal input			*Randomize*	Size of verbal input is not assessed in this dissertation;
Schemata			*Randomize*	Schemata are not assessed in this dissertation;
Contiguity	Spatial		*Fixed at:* One designated location of the display at the pilot's primary field of view;	One designated location may help reduce visual search;
	Temporal		*Fixed at:* Sequential	Sequential presentation of the V-T-T transcription can serve as: 1) reinforcement of the voice-only information presented first 2) means validation of content; and 3) a memory aid;
Modality	Visual-spatial		*Fixed at:* One visual-spatial task (e.g., monitoring, etc.)	Used to emulate the visual-spatial portion of the flying task;
	Verbal	Auditory	*Fixed at:* Voice	Used to emulate the two types of verbal display on a flight deck;
		Visual	*Fixed at:* Text	
Redundancy			*Suggested manipulation:* No redundancy (voice- only) vs.verbal redundancy (voice + V-T-T transcription)	Added redundancy in verbal display of information can reduce cognitive load;
Pace			*Fixed at: System-paced*	The pace of verbal communications in an operational environment such as a flight deck is dictated primarily by the phase of flight; Pilot has a very limited or no control of the rate, or timing of the incoming verbal information;

Factor	Experimental design	Hypothesis/Justification
Prior Knowledge/Level Expertise	*Suggested manipulation:* Low level of prior knowledge vs. high level of prior knowledge;	While prior knowledge and expertise reduce information processing limitations they also minimize the effects of redundancy manipulation effects. This factor may help determine what part of the pilot population would benefit the most from the implementation of V-T-T;
Verbal Ability	*Randomize*	Verbal ability is not assessed for this dissertation;
Spatial Ability	*Randomize*	Spatial ability is not assessed for this dissertation;
Interference	*Fixed at:* No articulatory suppression or irrelevant speech;	These types of interference effects manifest themselves when items-to-remember are presented visually; For this dissertation such items are presented auditorily;

CHAPTER THREE: EMPIRICAL EXAMINATION

Empirical studies in the aviation domain where pilot's cognitive load and the potential for miscommunications during controller-pilot voice radio communications was examined in the larger context of multimedia fashion information is presented in today's flight decks, and within a conceptual framework which reflected specifically the architecture of the phonological loop, are notably absent. Similarly, the utility of verbal redundancy as a design solution for managing pilot cognitive load was never examined in the context of the remaining role of voice in the future dual voice/data link communication environment. The goal of this dissertation is to assess the utility of a redundant (as compare to a non-redundant) verbal display of controller-pilot voice communications in limiting pilot's cognitive load to only the essential amount of processing resources required for understanding the information conveyed by the controller while at the same time reducing the potential for communication errors. In order to achieve this goal, three independent variables have been selected in terms of importance, practicality, and interest. Within the conceptual framework presented earlier in this dissertation, each factor and respective levels are described next.

Assessing the Effects of Verbal Display Redundancy on Cognitive Load and Performance

Verbal Display Redundancy Manipulation

Verbal display redundancy of the kind proposed in this dissertation brings to light the notion that pilot cognitive load and the potential for miscommunications between controllers and pilots can be managed by design. A voice-to-text transcription of the ATC clearances that is available for a review immediately following the delivery by voice (or at any time during the flight) may afford exactly that, and do so by providing the pilots with a redundant means to access, remember, and verify the content of controller's messages. One of the goals of this dissertation is to determine whether the relationship between redundancy, cognitive load, and performance, as established in the context of a learning environment, will be supported by empirical evidence from an operational environment such as a modern flight deck.

First factor's selection criteria

The first independent variable (IV1), Verbal Display Redundancy (i.e., redundant V-T-T vs. non-redundant "voice only" display of verbal information) was selected to isolate the extent to which verbal redundancy affected cognitive load and performance. Theoretically, redundancy effect (Sweller, van Merriënboer J. & Paas, 1998) occurs when different sources of information, which are intelligible in isolation (each source provides identical content only in a different form), are presented at the same time. Under these conditions, an unnecessary additional cognitive load is imposed by the very existence of multiple redundant sources of information. However, more often than not, the complexity of controller's messages is high and the intelligibility of voice communication in a flight deck is very poor. Therefore, it is possible that a redundant means for reliably accessing this information may actually have the opposite effect. In light of this, specific hypotheses are stated next.

Verbal Display Redundancy Hypotheses

Hypothesis 1

In a flight deck environment, a redundant display of verbal information originally delivered by voice will limit pilot's cognitive load to only the essential amount of processing resources required for understanding the information and correctly executing the required action.

Therefore, I hypothesized a main effect for verbal display redundancy. More specifically, I proposed (Figure 1):

H1A: Redundancy in the display of verbal information would be associated with lower cognitive load, and no redundancy with higher cognitive load.

H1B: Redundancy in the display of verbal information would be associated with greater performance, and no redundancy with lower performance.

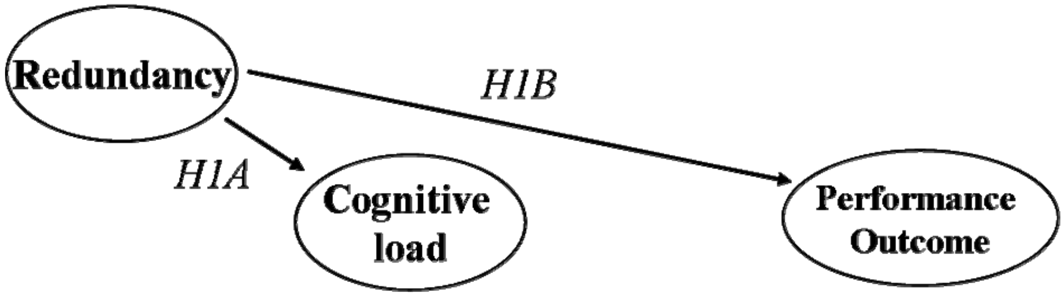


Figure 1 Hypothesis 1 - Redundancy → Cognitive Load & Performance

Assessing the Effects of Verbal Input Complexity on Cognitive Load and Performance

Verbal Information Complexity Manipulation

The most recent update of the cognitive load theory introduced the idea that intrinsic cognitive load (Chandler & Sweller, 1991; Pollock, Chandler & Sweller, 2002) depends on the level and complexity of information element interactivity. Element interactivity is low when each item can be understood, and learned without referencing any other items, and high when material cannot be fully understood until all of the elements, and their interactions are processed together (Paas, Renkl & Sweller, 2003). Once again, the goal here is to determine whether the relationship between complexity, cognitive load, and performance as identified in the context of a learning environment, will be supported by empirical evidence from an operational environment such as a modern flight deck.

Second factor's selection criteria

The second independent variable (IV2), Complexity of Verbal Information (i.e., low and high complexity verbal input) was selected to identify the extent to which complexity affected cognitive load and performance. In the context of controller-pilot communications, the level of complexity of an ATC message can be defined by the level of interactivity between the different elements (e.g., heading, altitude, etc.) of a clearance. Theoretically, a higher complexity clearance is associated with a higher cognitive load and higher propensity for communication errors.

Complexity of Verbal Input Hypotheses

Hypothesis 2

More processing resources are required to understand complex clearances because of the higher element interactivity. Therefore, I hypothesized a main effect for verbal input complexity (Figure 2).

H2A: Higher complexity clearances would be associated with higher cognitive load and lower complexity clearances with lower cognitive load.

H2B: Higher complexity clearances would be associated with lower performance and lower complexity clearances with greater performance.

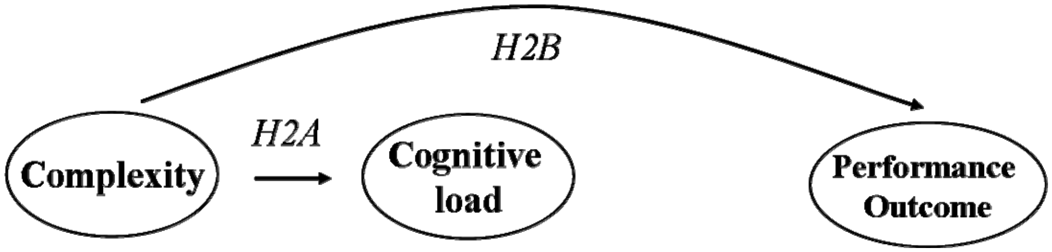


Figure 2 Hypothesis 2 - Complexity → Cognitive Load & Performance

Assessing the Effects of Level of Expertise on Cognitive Load and Performance

Level of Expertise Manipulation

Because of the major differences in the amount, and level of schemata automation, novices and experts process the same information very differently. Experts can access large amount of schemata and the availability of these knowledge structures can effectively remove the processing limitations of working memory and fundamentally change the characteristics of performance. These structures direct the allocation of cognitive resources and significantly influence the processing (Sweller, 2005; 2010; van Merriënboer & Sweller, 2005).

Third factor's selection criteria

The third independent variable (IV3), Level of Expertise (i.e., novices and experts) was selected to identify the extent to which level of expertise affected cognitive load and performance during controller-pilot voice communications. Level of expertise could be defined by the number of total flight hours, and pilot currency. Theoretically, a higher level of expertise is associated with a lower cognitive load and better performance during controller-pilot communications.

Level of Expertise Hypotheses

Hypothesis 3

I hypothesized a main effect for Level of expertise (Figure 3).

H3A: Higher level of expertise would be associated with lower cognitive load and lower level of expertise with higher cognitive load.

H3B: Higher level of expertise was expected to be associated with greater performance and lower level of expertise with lower performance.

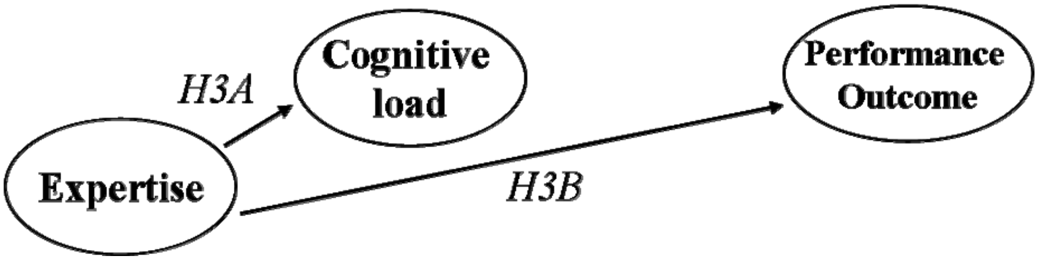


Figure 3 Hypothesis 3 Expertise → Cognitive Load & Performance

Assessing the Interaction of Display Redundancy, Complexity, and Level of Expertise on Cognitive Load and Performance

Hypothesis 4

I hypothesized a significant interaction between verbal display redundancy, verbal input complexity, and level of expertise on the measures of cognitive load and performance (Figure 4).

H4A: For the no redundancy/high complexity condition, cognitive load would be higher than for no redundancy/low complexity condition regardless of level of expertise.

H4B: For the redundancy/high complexity condition, cognitive load would be higher than for redundancy/low complexity condition regardless of level of expertise.

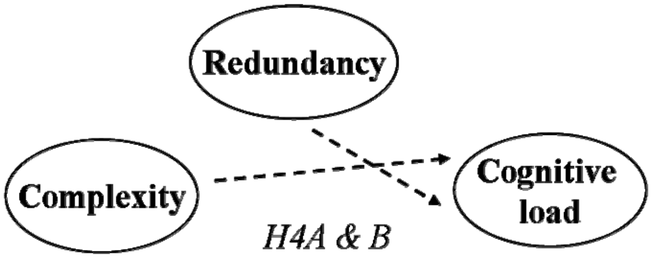


Figure 4 Hypotheses H4A and B - Complexity X Redundancy → Cognitive Load

H4C: In the redundancy/high complexity and redundancy/low complexity conditions, the differences in cognitive load between novices and experts would be significantly reduced compared to the differences between novices and experts in the no redundancy/high complexity and no redundancy/low complexity conditions (Figure 5 and Figure 6).

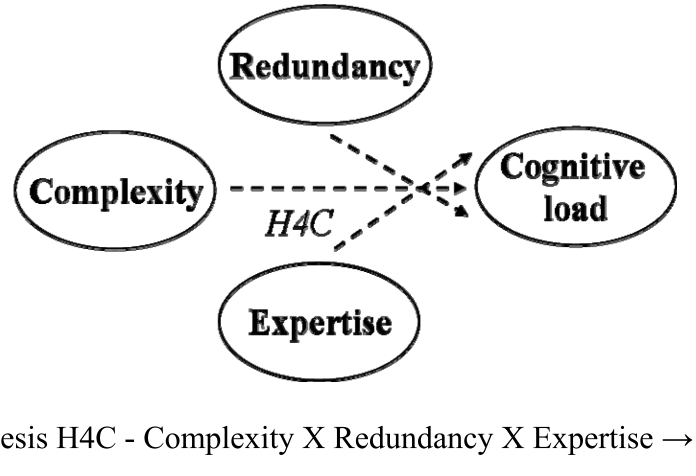


Figure 5 Hypothesis H4C - Complexity X Redundancy X Expertise → Cognitive Load

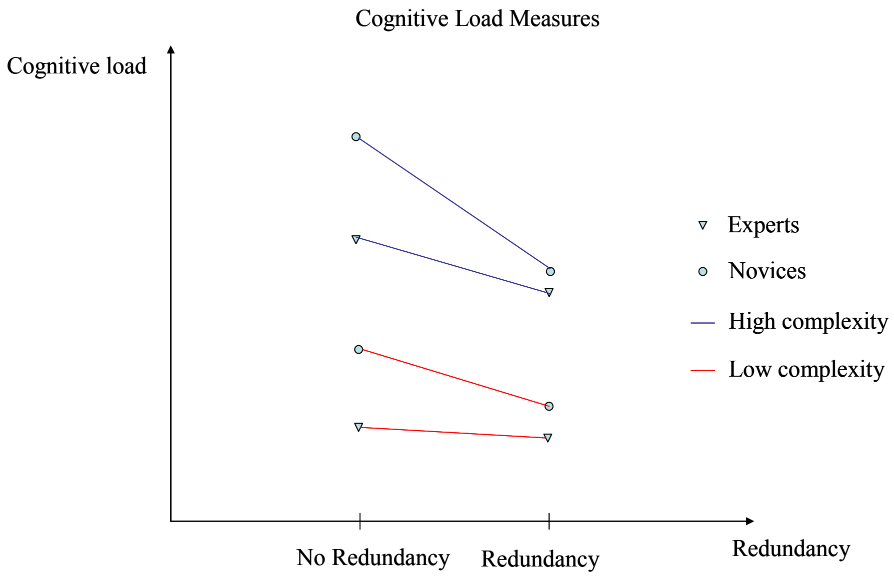


Figure 6 Hypothesized Complexity X Redundancy X Expertise interaction on the measures of cognitive load

Furthermore, I hypothesized a significant interaction between verbal display redundancy, verbal input complexity, and level of expertise on the measures of performance (on both primary and secondary task) (Figure 7).

H4D: For the no redundancy/high complexity condition, performance would be lower than for no redundancy/low complexity condition, regardless of level of expertise.

H4E: For the redundancy/high complexity condition, performance would be lower than for redundancy/low complexity condition, regardless of level of expertise.

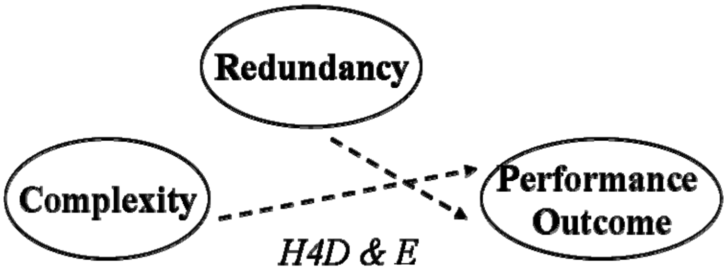


Figure 7 Hypotheses H4D and E - Complexity X Redundancy → Performance

H4F: For the redundancy/high complexity and redundancy/low complexity conditions, the differences in performance between novices and experts would be significantly reduced compared to the differences between novices and experts for the no redundancy/high complexity and no redundancy/low complexity conditions (Figure 8 and Figure 9.

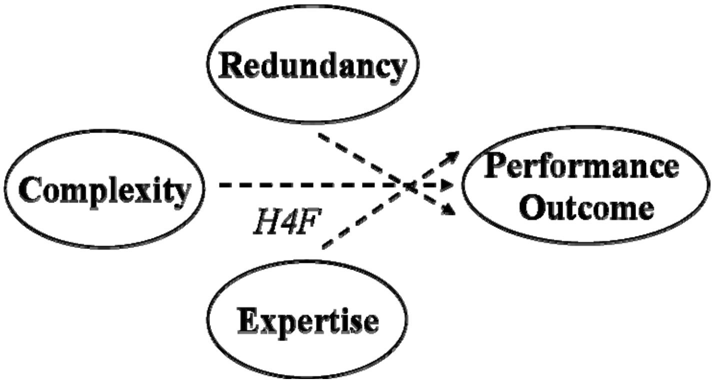


Figure 8 Hypothesis H4F - Complexity X Redundancy X Expertise → Performance

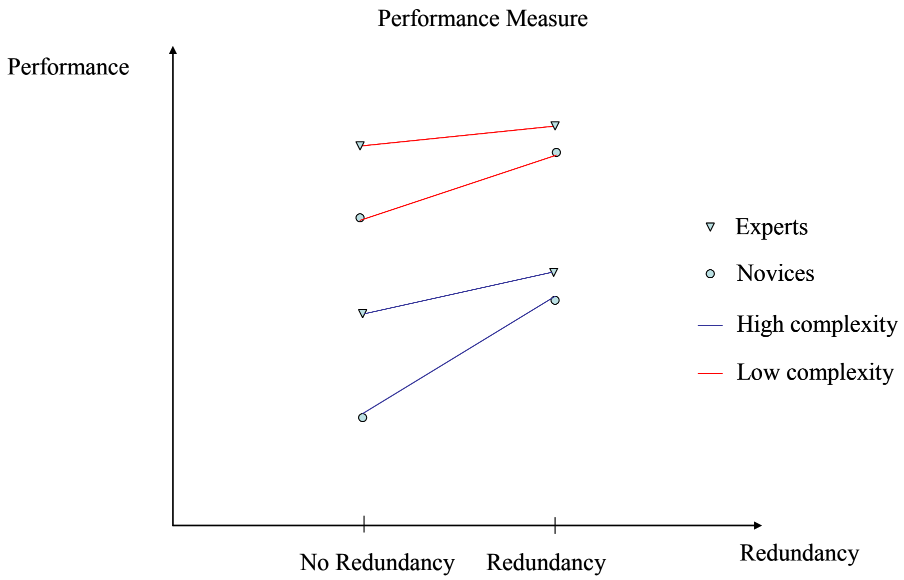


Figure 9 Hypothesized Complexity X Redundancy X Expertise interaction on the measures of performance

Assessing the Relationship between Manipulated Factors, Cognitive Load, and Performance

The mediating role of cognitive load in the relationship between presentation format, complexity of learning material, and learning outcomes, has been well documented in the instructional research literature (Paas & van Merrienboer, 1994). Theoretically, it was important to verify whether cognitive load would play the same or a similar role in the relationship between the same variables in the context of an operational environment such as a modern flight deck which made it particularly important because the theoretical models employed in this dissertation originated from instructional research domain. This dissertation manipulated complexity, redundancy, and level of expertise, which were expected to differentially affect cognitive load and performance, as hypothesized above. Furthermore, in order to include cognitive load as a construct in the pilot-system interface design models, it was essential to validate its role as a mediator between pilot-system interface characteristics, and performance outcomes. Therefore, moderated mediation was used to examine whether cognitive load mediated the relationship between the interaction of complexity, redundancy, and expertise in predicting performance for a variety of pilot populations. It was essential to verify that managing cognitive load by design was critical to the performance outcome, as well as, that the pilot-system interface is designed such that it has no adverse effects on performance across the different levels of expertise.

Hypothesis 5

H5: The Redundancy X Complexity X Expertise interaction and performance outcome would be mediated by cognitive load (Figure 10).

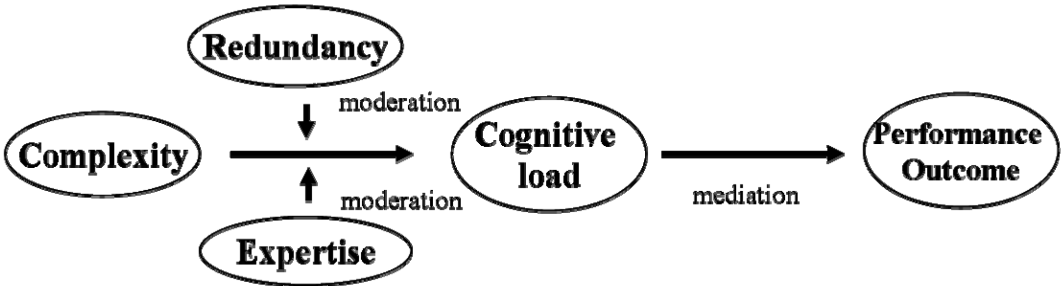


Figure 10 Representation of the moderated mediation model.

CHAPTER FOUR: METHOD

Participants

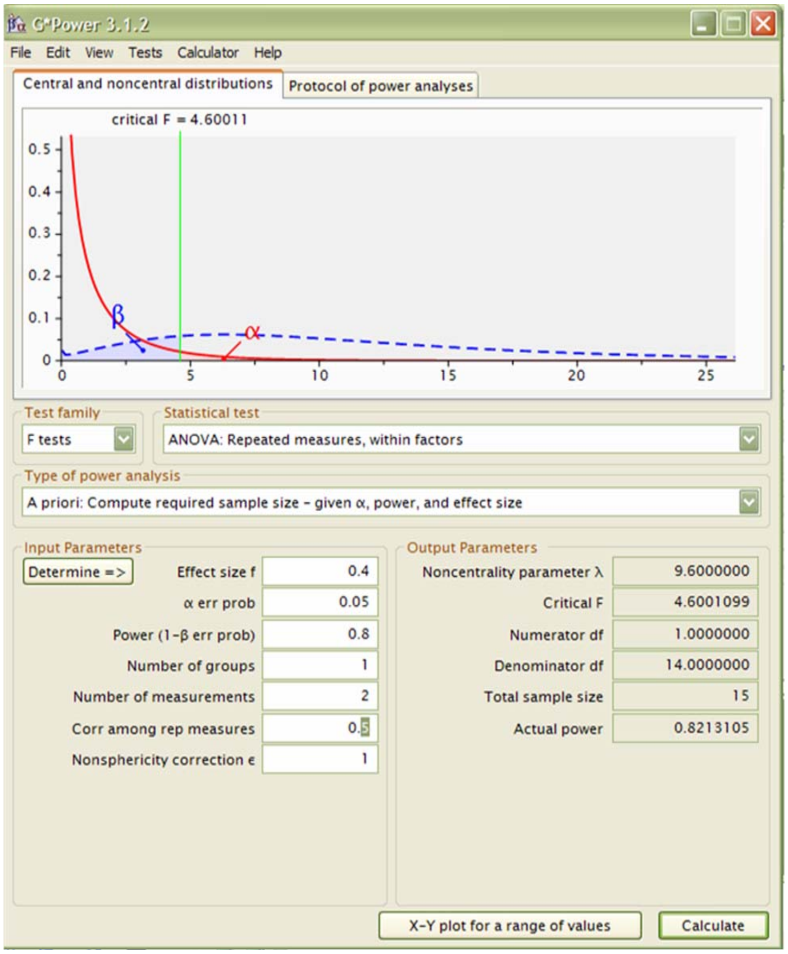
Seventeen student pilots (i.e., “novices”) and seventeen instructor pilots (i.e., “experts”) were recruited to participate in this dissertation research. The inclusion criteria for “novices” consisted of the following:

- Holder of at least a Private Pilot License;
- To have flown in the last 60 days;
- To have between 100 and 250 total flight hours;

The inclusion criteria for “experts” consisted of the following:

- Holder of at least an Instrument Pilot License;
- To have flown in the last 60 days;
- To have at least 500 total flight hours;

A power analysis was conducted using G*Power 3.1 (Faul, Erdfelder, Lamg, & Buchner, 2007) to determine the number of required participants according to a specified affect size and overall power. Three types of a priori assessment were performed, (a) ANOVA: Repeated measures, within factors (Figure 11); (b) ANOVA: Repeated measures, between factors (Figure 12); and (c) ANOVA: Repeated measures, within-between interaction (Figure 13). Cohen (1988) defines *f* s of 0.1, 0.25, and 0.4 as small, medium, and large effect size. For the a priori assessments conducted for this dissertation research, an effect size of 0.4 was entered. Also, a power level of .80 was adopted, which is an acceptable compromise between high and low power (Cohen, 1988). A resulting sample of 34 participants is needed based on these analyses.



Test family: F tests

Statistical test: ANOVA: Repeated measures, within factors

Type of power analysis: A priori: Compute required sample size – given α, power, and effect size

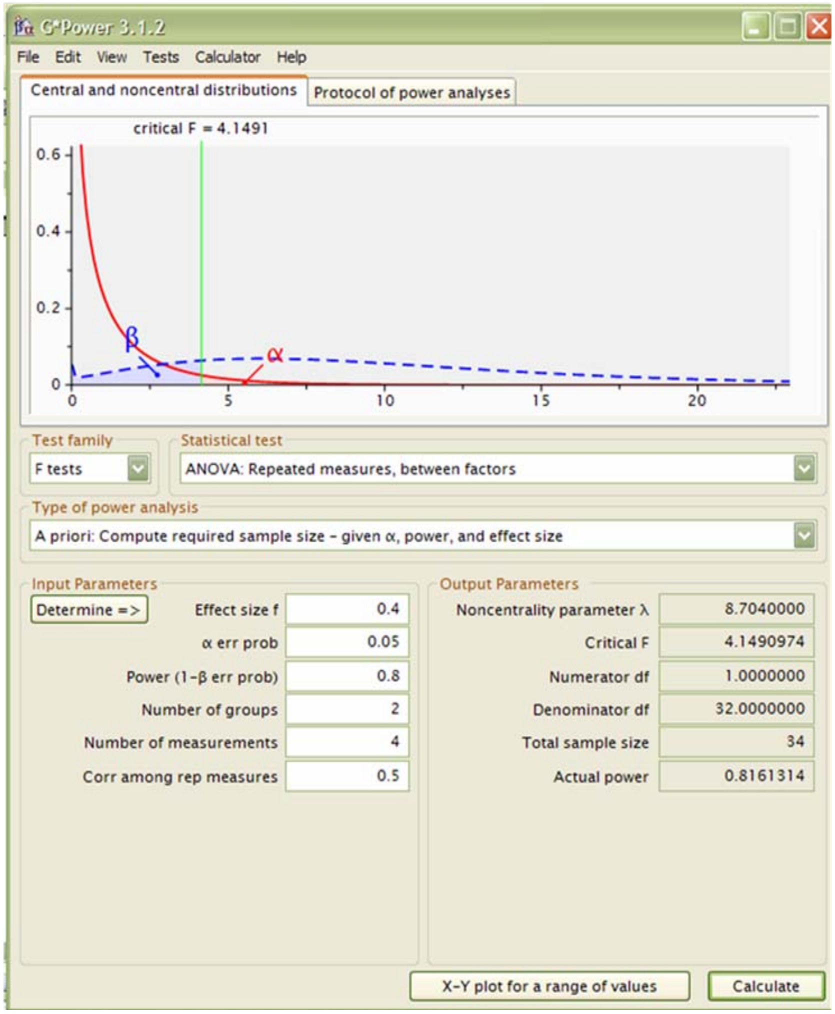
Input Parameters

Effect size f	0.4
α err prob	0.05
Power (1-β err prob)	0.8
Number of groups	1
Number of measurements	2
Corr among rep measures	0.5
Nonsphericity correction ε	1

Output Parameters

Noncentrality parameter λ	9.6000000
Critical F	4.6001099
Numerator df	1.0000000
Denominator df	14.0000000
Total sample size	15
Actual power	0.8213105

Figure 11 Screenshot of G*Power for ANOVA: Repeated measures, within factors.



Test family: F tests

Statistical test: ANOVA: Repeated measures, between factors

Type of power analysis: A priori: Compute required sample size – given α, power, and effect size

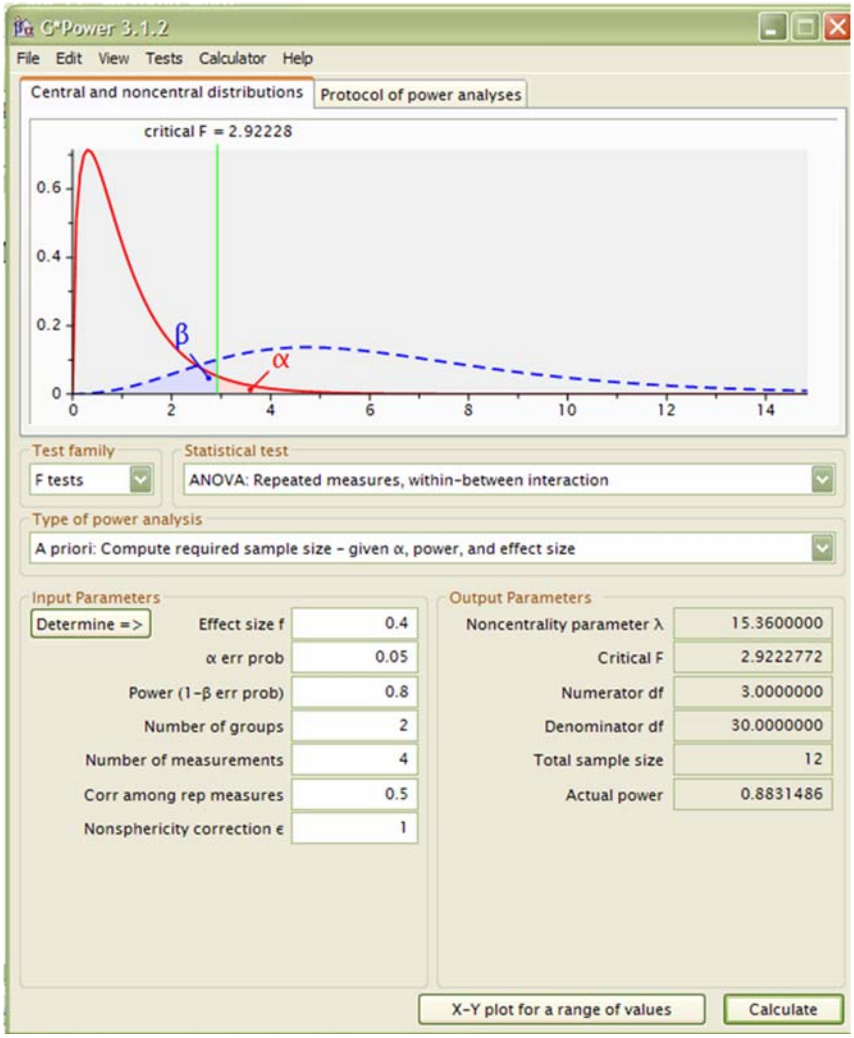
Input Parameters

Effect size f: 0.4
α err prob: 0.05
Power (1-β err prob): 0.8
Number of groups: 2
Number of measurements: 4
Corr among rep measures: 0.5

Output Parameters

Noncentrality parameter λ: 8.7040000
Critical F: 4.1490974
Numerator df: 1.0000000
Denominator df: 32.0000000
Total sample size: 34
Actual power: 0.8161314

Figure 12 Screenshot of G*Power for ANOVA: Repeated measures, between factors.



Test family: F tests

Statistical test: ANOVA: Repeated measures, within-between interaction

Type of power analysis: A priori: Compute required sample size – given α, power, and effect size

Input Parameters	
Effect size f	0.4
α err prob	0.05
Power (1-β err prob)	0.8
Number of groups	2
Number of measurements	4
Corr among rep measures	0.5
Nonsphericity correction ε	1

Output Parameters	
Noncentrality parameter λ	15.3600000
Critical F	2.9222772
Numerator df	3.0000000
Denominator df	30.0000000
Total sample size	12
Actual power	0.8831486

Figure 13 Screenshot of G*Power for ANOVA: Repeated measures, within-between interaction.

Design

A 2 Verbal Display Redundancy (no-redundancy and redundancy) X 2 Verbal Input Complexity (low and high) X 2 Level of Expertise (novices and experts) mixed-model design (Table 7) was used for the study. General Linear Model Repeated Measures analysis of variance and Hierarchical Linear Model analyses were used for statistical tests on the cognitive load and performance measures.

Table 7 Matrix of the Experimental Conditions

	Novices		Experts	
	No Redundancy	Redundancy	No Redundancy	Redundancy
Low Verbal Input Complexity	Low complexity clearances without V-T-T transcription	Low complexity clearances with V-T-T transcription	Low complexity clearances without V-T-T transcription	Low complexity clearances with V-T-T transcription
High Verbal Input Complexity	High complexity clearances without V-T-T transcription	High complexity clearances with V-T-T transcription	High complexity clearances without V-T-T transcription	High complexity clearances with V-T-T transcription

Materials and Apparatus

Instrument Flight Rules (IFR) domain was adopted for this dissertation. The low and high complexity clearances were selected from the Jeppesen *ATC Clears* IFR Clearance Shorthand instructional CD. A V-T-T transcription of these clearances was displayed in its entirety (as opposed to a scrollable presentation) on a V-T-T widget. The shorthand abbreviations and symbols utilized for the display of V-T-T transcription were the same as those used in the *ATC Clears* instructional CD. When the V-T-T transcription was not shown (no redundancy condition) the V-T-T widget was blank. X-Plane® desktop simulation software (Figure 14) was used to emulate the ongoing visual-spatial task (e.g., monitoring).

C 64V ABI (T) V17 BPR V66 M80 EX HIGHER < BPR CT FT.W CTR 127.0 < 40 SQ 3516 VOID 1227

Figure 14 Screen shot of X-Plane® desktop simulation used in the study

The order of each test sequence (one per participant) of 20 clearances (10 with low, and 10 with high level of complexity) was randomized. The time interval between any two clearances was also randomized and varied in length between 45 sec and 2 min. Fifty percent of each sequence included a V-T-T transcription where 5 clearances were with low level of complexity, and 5 with high level of complexity (Table 8).

Table 8 Number of stimuli per participant per condition

	No Redundancy	Redundancy
Low Verbal Input Complexity	5 Low complexity clearances without V-T-T transcription	5 Low complexity clearances with V-T-T transcription
High Verbal Input Complexity	5 High complexity clearances without V-T-T transcription	5 High complexity clearances with V-T-T transcription

The following materials were available for use by the participants for the duration of the tests:

- A laminated job-aid with the shorthand abbreviations and symbols used to display the clearances on the V-T-T transcription widget
- A list of identifiers for the starting and ending points of all 20 clearances used in the experiment
- Dallas-Fort Worth area aeronautical charts
- NASA TLX rating scale definitions
- NASA TLX participant instructions (for rating and sources of workload evaluation)
- NASA TLX assessment materials
- Usability survey
- Notepad
- Pen and pencil.

All necessary sound editing was conducted using Sony Sound Forge® Audio Studio 10. Participants wore headsets with a microphone. A digital recording device was used for capturing the clearance read-backs. Participants were given a score sheet to manually record the time in minutes and seconds the green annunciator light was ON.

Task

Voice communication tasks carried by a pilot usually include receiving, processing, and acting upon instructions issued by a controller therefore requiring the pilot to retain the information in his/her memory for a short period of time between the receiving and acting upon these instructions. The controller's messages usually contain more than one instruction (e.g., heading, altitude, and contact frequency). Frequently, the pilot is also engaged in some kind of a "distracting" (with respect to the action required by the controller's instructions) task such as, for example performing a checklist (Loftus, Dark & Williams, 1979). In a Brown-Peterson research paradigm, participants are required to perform a very similar task (Brown, 1958; Peterson & Peterson, 1959). In the original studies the experimenter first read aloud a consonant trigram (e.g., BDF) followed by a three-digit number. Then, to prevent rehearsal of the trigram, the participants were asked to count backwards (distractor task) from the three-digit number, by three or four, for a certain period of time. At the end of this period, the participant was expected to recall the three consonants in order. Peterson and Peterson (1959) varied the time period participants counted backwards. The results showed that the proportion of consonants correctly recalled was a function of the duration of the distractor task. After counting backwards for a period as short as 18 sec, the performance declined to some asymptotic level of only about 10% correctly recalled items.

Flying is a mostly visual-spatial task, and as such, it requires the integration of information that is visually based, and spatial in nature. It also entails a strict task priority hierarchy (i.e., “Aviate, Navigate, Communicate”) where voice communications have a mostly supporting role, and traditionally, in two-pilot operations, the pilot monitoring is in charge of this task. Also, as previously discussed, pilots shorthand the clearances they receive from the ATC, and to do so by using a specific order or format. They are expected (though not required) to read back the parts of any clearance containing altitude assignments, radar vectors, or any portion of the clearance requesting verification (ATC may request a read-back when certain factors such as the complexity of the clearance suggest a need).

To emulate the tasks performed by a flight crew, and more specifically, the tasks conducted by the pilot monitoring, for the purpose of this dissertation two research paradigms were employed. First, dual-task methodology was utilized where the primary task was to listen and read back ATC clearances (verbal) and the secondary task was to monitor a cockpit indicator light (visual). Second, within the primary task, a Brown-Peterson methodology was employed. Specifically, one IFR clearance playback was presented at a time. Although in reality, pilots can request a clearance to be repeated (e.g., “Say again”), the clearances during the test portion of this experiment were not repeated. When a clearance playback ended, the participants were asked to start counting out loud, backwards from a randomly generated 3 digit number, by three, for a period of 20 seconds (i.e., distractor task) (Peterson & Peterson, 1959). After completion of the “distractor” task, participants were asked to read back the clearance in its entirety (it was up to each individual to use their own shorthand notes or the V-T-T, if available). At the end of each read-back, the participants were prompted to fill out the NASA TLX questionnaire. For the secondary task, the participants were asked to continuously monitor a green annunciator light on the upper instrument panel of the cockpit in the X-Plane® simulation and write down the times the light was ON by recording the time in minutes and seconds from the digital clock provided. The light was ON at random intervals of time varying between 2 and 3 min. The secondary task was performed continuously through out the test trials.

The primary task was a basic verbal communication task (listening and responding/reading back a clearance) and the secondary task was a basic visual task (monitoring). Both tasks resembled very closely the tasks performed by the pilot monitoring in a two-pilot flight crew. The training session prior to the actual data collection, as well as, the materials (e.g., area aeronautical charts) provided to the participants for use during the test trials, ensured the successful completion of their participation in the study.

Cognitive Load Assessment

According to Paas and van Merriënboer (1994), as a construct, cognitive load contains three measurable dimensions reflecting mental load, mental effort, and performance. The authors define the aspect of cognitive load, indicative of the estimated demand on cognitive capacity, and originating from the interaction between the attributes of the task, and the characteristics of the individual performing the task, as mental load. Following to the same model, the dimension of cognitive load associated with the actual cognitive capacity allocated to sustain the demands imposed by the task, and considered to reflect the actual cognitive load, is mental effort. Importantly, this facet of cognitive load can be measured while participants are performing a task. The third aspect of cognitive load - performance - can be measured in terms the number of errors, number of correct test items, or time on task. Per Paas and van Merriënboer (1994b), however, the estimates of mental effort may not necessarily be reflected in mental load and performance measures. That is, equal performance levels may not be achieved by the same amount of effort.

A wide variety of analytical and empirical methods has been used to measure the different aspects of cognitive load (Xie & Salvendy, 2000). Analytical methods are focused at estimating mental load, and use subject matter experts' assessment, analytical data derived by employing task modeling, and task analysis techniques. Alternatively, empirical methods measure mental effort and performance by gathering subjective data using rating scales, performance data by utilizing dual-task methods, physiological, as well as, neuroimaging techniques. In cognitive load research as a whole, however, subjective rating scales ((Paas, Renkl & Sweller, 2003; Kalyuga, Chandler & Sweller, 1999) and dual-task techniques (Chandler & Sweller, 1996; Marcus, Cooper, & Sweller, 1996) have been used most frequently. Subjective methods usually involve a questionnaire with one or multiple differential rating scales where the participants can indicate the level of actual cognitive load they experienced. Rating scales are based on the assumption that participants are able to assess, and report the amount of mental effort they have expended on a particular task (Gopher & Braune, 1984). Although this frequently used technique (Paas et al., 2003) appears to be able to reliably assess the subjective perception of invested effort, there is some ambiguity about how exactly mental effort relates to actual cognitive load (Brunken, Plass & Leutner, 2003). Kalyuga, Chandler and Sweller (1999) used ratings of the difficulty rather than ratings of mental effort, and reported a high sensitivity of these scales in identifying differences in training approach.

Objective measures of cognitive load based on task performance, are frequently used in a dual-task paradigm, and are closely related to cognitive load in working memory research

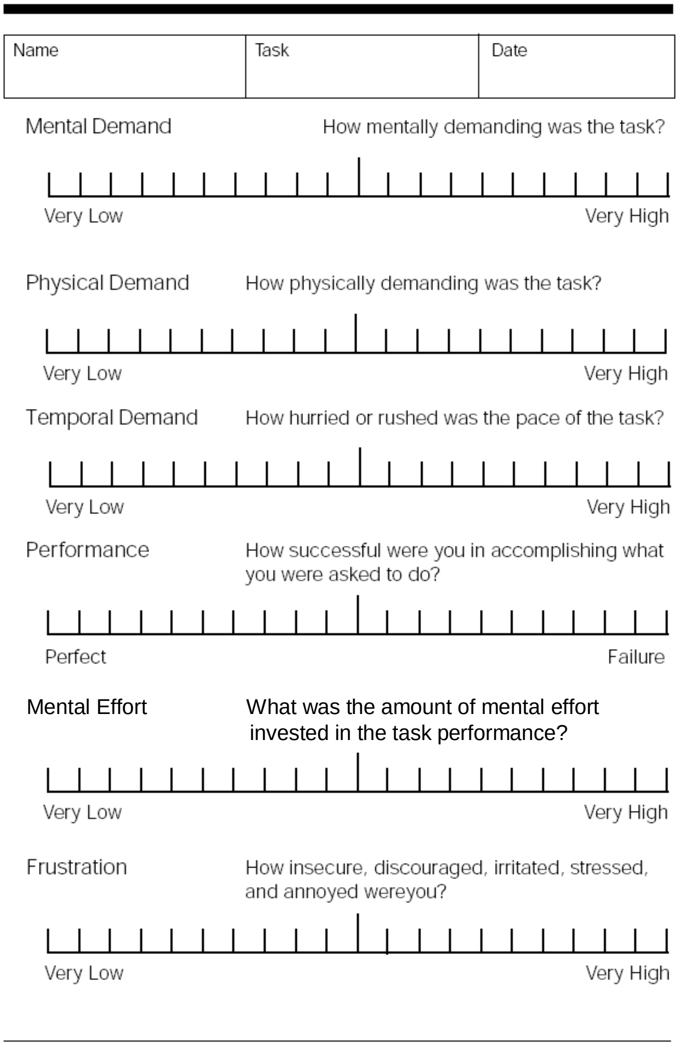
(Baddeley, 1986). Dual-task method assumes limited cognitive resources that can be dynamically allocated to different aspects of a task. Therefore, if two tasks were to be performed simultaneously, and if both tasks require the same resources in verbal and/or visual working memory, the available verbal and visual resources have to be distributed between these two tasks. Two approaches may be applied in a dual-task paradigm. One approach is to add a secondary task with the only intent of introducing memory load. The dependent variable of interest then is the primary task performance, which is expected to degrade in a dual-task condition compared to a single-task condition (i.e., the primary task alone). Another approach is to use secondary task performance as a measure of the memory load induced by the primary task. In this case, the performance on the secondary task is the variable of interest. If different versions of a primary task induce different amounts of memory load, then the performance in the secondary task should vary accordingly. Primary task and secondary task measurements include error rate, reaction time, accuracy, etc.

While subjective rating scales of mental effort can only be reasonably applied after the task execution, dual-task techniques make it possible to measure cognitive load at the point in time when the load is introduced (as the primary and secondary tasks are performed to at the same time). In addition, based on working memory research, there are different secondary tasks that are linked to different stages of human information processing (HIP) (Baddeley, 1986). These tasks can help identify the stage of HIP where cognitive load is imposed. Dual-task paradigm works well in within-subjects designs (as compare to between-subjects designs) because it allows the measures of cognitive load to be independent from individual differences, such as spatial and verbal abilities, or prior knowledge. In the context of using secondary task only as a memory load, in order for it to be successful, it has to be set up to continuously expend all of the available “free cognitive capacity” per Brunken, Plass, and Leutner (2003). When the difference between the total cognitive load and the processing capacity of the visual or auditory working memory is minimal or zero, the cognitive load is high and that difference may be used as a basis for direct measurement of cognitive load. Lastly, it should be pointed out that while the cognitive load theory makes a distinction between intrinsic, extraneous, and germane load, researchers have only been able to measure the total cognitive load, and not any of its three components (Paas, Tuovinen, Tabbers & van Gerven, 2003).

Cognitive Load Measure

To achieve the ultimate system performance goals, system designers need to account for the overall operator workload at all stages of system design and operation. The NASA Task Load Index (TLX) is a multi-dimensional subjective workload rating technique, which integrates the weighted subjective responses driven by perceptions of task demand (Hart & Staveland, 1988). It was developed based on the assumption that a combination of six dimensions (mental demand, physical demand, temporal demand, perceived performance, effort, and frustration level) represent the “workload” experienced by most people performing most tasks. These dimensions were selected after an extensive analysis of factors that identify the subjective experience of workload for different people performing activities ranging from simple to complex tasks such as flying an aircraft (Rehman, 1995). Detailed description of the development process and theoretical rationale behind the NASA TLX scale are presented in a chapter published in 1988 by Hart & Staveland.

According to Hart (2006), most of the studies, which used NASA TLX, addressed a question about interface design and 31% of them focused on visual and/or auditory displays. Furthermore, the author reported that a common variation of the scale is to conduct subscale-rating analyses instead of generating a single overall workload score. Over 40 studies used this approach and demonstrated the potency of the scale and the diagnostic value of the component subscales (Hart, 2006). The high reliability, sensitivity, and utility of the NASA TLX component ratings allow designers to very narrowly identify sources of a workload or performance problem. As the focus of this dissertation research was primarily on the cognitive load (mental demand and mental effort) during controller-pilot voice communications, a similar approach was applied here, as well. The NASA TLX scale was conducted in its entirety. In addition, to focus specifically on the perceived amount of mental effort invested in the performance of the task, the description of the NASA TLX "Effort" subscale was modified as shown on Figure 15.



Name	Task	Date

Mental Demand — How mentally demanding was the task?

Very Low — Very High

Physical Demand — How physically demanding was the task?

Very Low — Very High

Temporal Demand — How hurried or rushed was the pace of the task?

Very Low — Very High

Performance — How successful were you in accomplishing what you were asked to do?

Perfect — Failure

Mental Effort — What was the amount of mental effort invested in the task performance?

Very Low — Very High

Frustration — How insecure, discouraged, irritated, stressed, and annoyed wereyou?

Very Low — Very High

Figure 15 NASA TLX with modified “Effort” scale.

Primary Task Performance Measures

Within the conceptual framework of this dissertation, the average percent correct read-back from 5 clearances per condition was used as a measure of primary task performance. The total of 20 test clearance read-backs per participant were scored using the respective audio recording by calculating the percent correct. More specifically, for each clearance, the number of correctly read back items was calculated (in percent correct) based on the weighted scoring system where, (a) all altitudes, destinations, holding patterns, routing, and “Expect further clearance” items were considered a priority “one”, and were given a weight of 3; (b) all frequencies and squawks were considered a priority “two” and given a weight of 2; and (c) any other information (unless associated with safety of flight, e.g., bad weather) was considered a priority “three”, and given a weight of one. Each clearance was divided into self-contained chunks of information, or items, representing one of the categories specified above. In the scoring process, when only a portion of an item was read back correctly, partial credit was given. For example, if a participant read back only the first half of a self-contained item such as “Maintain 14, 000 and advise” (Table 9) and omitted “and advise”, a score of 1.5 instead of 3, was given for this item. Furthermore, a weight of “3” was assigned to each component of a one bad weather item in clearance #15 (

Table 10). For each clearance, an ideal total score was derived by adding all the weights. Finally, using the actual scores, a percent correct clearance read-back was calculated.

Table 9 Example of a clearance with the weighted scoring system used to calculate percent correctly read back clearances.

ATC clears (call sign) to the Meacham Airport	via V18	Maintain VFR on top	If not VFR on top at 14,000	Maintain 14,000 and advise	No top reports available	Contact Fort Worth Center on 127.6	Squawk 1422
3	3	3	3	3	1	2	2

Table 10 Example of clearance with weather information weighted priority "one" due to safety of flight implications

(Call sign) cleared to the Dallas Love Airport	Direct Blue Ridge VORTAC	Descend and maintain 1-2,000 12,000	Report passing 1-5,000, 15,000	Depart Blue Ridge VORTAC	For vectors to runway 31 right	Heading 210	ILS final approach course	Landing runway 31 right	Dallas Love weather, measured ceiling 600 overcast	Visibility 2 mi	Light rain showers	Temp 52	Wind 290 at 4	Altimeter 3013
3	3	3	3	3	3	3	3	3	3	3	3	3	3	3

Secondary Task Performance Measures

For the duration of each test trial session, an automatic count of the number of times the green annunciator light is ON with a time stamp was recorded via a software program built in the X-Plane® simulation. The total number of times the light was ON as automatically captured by the computer program compared to the number recorded manually by each participant in percent correct was planned for use as a measure of secondary task performance. Due to a software limitation however, the automatically captured data only reflected that total number of times the annunciator light was ON during each test sequence without the ability to allocate these data to each experimental condition. Consequently, no analyses were conducted on these data.

Usability Measure

In addition to the cognitive load, and primary and secondary tasks performance assessment, a survey of the usability of V-T-T transcription display was conducted to gauge participants' perception of the ease of use, ease of interpretation, usefulness, overall location, and layout of the V-T-T transcription widget.

A summary of the measures collected per participant and experimental condition is presented in Table 11 below.

Table 11 Summary of measures per participant and experimental condition

	No Redundancy	Redundancy
Low Verbal Input Complexity	• The average of 5 total NASA TLX scores; • The average of 5 scores for each of the 6 NASA TLX subscales; • The average score of 5 clearance read-backs (primary task) in % correct; • Overall % correct score for annunciator light monitoring (secondary task);	• The average of 5 total NASA TLX scores; • The average of 5 scores for each of the 6 NASA TLX subscales; • The average score of 5 clearance read-backs (primary task) in % correct; • Overall % correct score for annunciator light monitoring (secondary task);
High Verbal Input Complexity	• The average of 5 total NASA TLX scores; • The average of 5 scores for each of the 6 NASA TLX subscales; • The average score of 5 clearance read-backs (primary task) in % correct; • Overall % correct score for annunciator light monitoring (secondary task);	• The average of 5 total NASA TLX scores; • The average of 5 scores for each of the 6 NASA TLX subscales; • The average score of 5 clearance read-backs (primary task) in % correct; • Overall % correct score for annunciator light monitoring (secondary task);
	One Sources-of-Workload set of weights; One usability survey results;	

Procedures

Prior to conducting any portion of the experiment, all participants were required to read and sign an informed consent form (Appendix B). Participants were then briefed on the purpose of the experiment and asked to fill out a demographics form. The experiment was conducted one participant at a time.

All the clearances selected for this study were sample clearances for IFR departures and arrivals at airports in the Dallas-Fort Worth area, therefore, after the briefing, the participants were given time to study the appropriate area aeronautical charts. A laminated print out with the shorthand symbols and abbreviations was available for reference during both the training and test trials sessions in case any of the participants are either not very familiar with that particular method, or have been using an alternative shorthand. After the participants have studied the charts, and the shorthand print out, they were instructed on how to use the equipment, and received an approximately 30 min of training including trial runs, representative of all experimental conditions, in a random order. The participants were then asked to begin the test trials. The NASA TLX workload ratings survey (Appendix D) was administered after each clearance read-back was completed. The entire test session was audio taped. Through out the test session, the participants were asked to manually record the time in minutes and seconds (from a digital clock) when the green annunciator light is ON. A score sheet with these times was collected at the end of the test. The NASA TLX Sources-of-Workload evaluation was conducted after the test trials were complete. The V-T-T transcription usability survey (Appendix E) was administered last. At the conclusion of the post-test surveys and evaluations, the participants were debriefed and dismissed.

CHAPTER FIVE: RESULTS

For this study, based on the hypothesized relationships between variables, analyses consisted of a series of mixed-model ANOVAs using SPSS General Linear Model Repeated Measures, as well as, Hierarchical Linear Model (HLM) using SPSS Mixed Models - Linear. Descriptive statistics are presented first, and what follows is a more detailed description of the analyses for each hypothesis.

Data Screening

Data collected during the experiment was screened for outliers, and normality of the dependent variable (DV) measures was assessed. The skewness and kurtosis of the repeated measures satisfied the assumption of normality. The assumption of homogeneity of variance was supported by Levene’s test for equality of variance.

Descriptive data

The population of participants was equally divided into novices and experts based on the criteria specified in section Participants of this dissertation. Five participants from the novices’ group and eight from the experts’ group reported having English as a second language (Table 12).

Table 12 Population frequency per level of expertise and English as a second language

	Novices	Experts	Overall L2	Overall N
Level of Expertise	17	17		34
English as a Second Language (L2)	5	8	13	

Inter-correlations, means, and standard deviations between important variables are outlined in Table 13

Table 13 Inter-correlations, means, and standard deviations for total raw NASA TLX, Mental Demand (MD), and Physical Demand (PD) scores.

.	Total Raw TLX LC-No R	Total Raw TLX HC-No R	Total Raw TLX LC-Yes R	Total Raw TLX HC-Yes R	Raw TLX MD LC-No R	Raw TLX MD HC-No R	Raw TLX MD LC-Yes R	Raw TLX MD HC-Yes R	Raw TLX PD LC-No R	Raw TLX PD HC-No R	Raw TLX PD LC-Yes R	Raw TLX PD HC-Yes R
Total Raw TLX Low Complexity No Redundancy	-											
Total Raw TLX High Complexity No Redundancy	.627**	-										
Total Raw TLX Low Complexity Yes Redundancy	.784**	.726**	-									
Total Raw TLX High Complexity Yes Redundancy	.755**	.622**	.846**	-								
Raw TLX Mental Demand Low Complexity No Redundancy	.829**	.736**	.657**	.621**	-							
Raw TLX Mental Demand High Complexity No Redundancy	.567**	.864**	.638**	.541**	.832**	-						
Raw TLX Mental Demand Low Complexity Yes Redundancy	.615**	.706**	.784**	.660**	.761**	.798**	-					
Raw TLX Mental Demand High Complexity Yes Redundancy	.671**	.721**	.732**	.793**	.830**	.806**	.876**	-				
Raw TLX Physical Demand Low Complexity No Redundancy	.275	.101	.365*	.391*	.178	.077	.116	.227	-			
Raw TLX Physical Demand High Complexity No Redundancy	.276	.105	.357*	.380*	.180	.083	.106	.212	.997**	-		
Raw TLX Physical Demand Low Complexity Yes Redundancy	.274	.107	.399*	.418*	.159	.063	.149	.247	.983**	.966**	-	
Raw TLX Physical Demand High Complexity Yes Redundancy	.276	.101	.376*	.405*	.176	.075	.132	.245	.996**	.985**	.994**	-

**p<0.01
*p<0.05
N=34

.	Total Raw TLX LC-No R	Total Raw TLX HC-No R	Total Raw TLX LC-Yes R	Total Raw TLX HC-Yes R	Raw TLX MD LC-No R	Raw TLX MD HC-No R	Raw TLX MD LC-Yes R	Raw TLX MD HC-Yes R	Raw TLX PD LC-No R	Raw TLX PD HC-No R	Raw TLX PD LC-Yes R	Raw TLX PD HC-Yes R
Raw Temporal Demand Score Low No R	.652**	.347*	.436**	.442**	.478**	.236	.348*	.387*	.201	.193	.220	.215
Raw Temporal Demand Score High No R	.309	.607**	.351*	.346*	.354*	.461**	.355*	.395*	.047	.040	.068	.060
Raw Temporal Demand Score Low Yes R	.487**	.543**	.634**	.532**	.388*	.403*	.534**	.515**	.201	.188	.236	.219
Raw Temporal Demand Score High Yes R	.480**	.486**	.518**	.671**	.413*	.356*	.497**	.611**	.155	.139	.190	.176
Raw Performance Score Low No R	.771**	.473**	.485**	.428*	.538**	.335	.296	.292	.026	.042	.001	.012
Raw Performance Score High No R	.342*	.770**	.423*	.378*	.433*	.620**	.458**	.428*	-.080	-.068	-.085	-.086
Raw Performance Score Low Yes R	.380*	.219	.483**	.273	.017	.007	.063	-.070	.230	.246	.225	.216
Raw Performance Score High Yes R	.287	.105	.338	.500**	-.051	-.041	-.023	-.007	.199	.220	.173	.180
Raw Mental Effort Score Low No R	.814**	.625**	.658**	.647**	.903**	.741**	.768**	.817**	.171	.164	.176	.179
Raw Mental Effort Score High No R	.472**	.879**	.589**	.494**	.764**	.923**	.776**	.765**	-.021	-.022	-.007	-.014
Raw Mental Effort Score Low Yes R	.663**	.603**	.807**	.752**	.736**	.708**	.776**	.793**	.250	.242	.268	.261
Raw Mental Effort Score High Yes R	.581**	.502**	.694**	.805**	.658**	.645**	.694**	.790**	.286	.268	.319	.308

**$p<0.01$
*$p<0.05$
N=34

	Raw TLX TD LC- No R	Raw TLX TD HC- No R	Raw TLX TD LC- Yes R	Raw TLX TD HC- Yes R	Raw TLX P LC- No R	Raw TLX P HC- No R	Raw TLX P LC- Yes R	Raw TLX P HC- Yes R	Raw TLX ME LC- No R	Raw TLX ME HC- No R	Raw TLX ME LC- Yes R	Raw TLX ME HC- Yes R
Raw TLX Temporal Demand Low Complexity No Redundancy	-											
Raw TLX Temporal Demand High Complexity No Redundancy	.657**	-										
Raw TLX Temporal Demand Low Complexity Yes Redundancy	.698**	.680**	-									
Raw TLX Temporal Demand High Complexity Yes Redundancy	.713**	.774**	.766**	-								
Raw TLX Performance Low Complexity No Redundancy	.481**	.187	.293	.271	-							
Raw TLX Performance High Complexity No Redundancy	.050	.364*	.184	.240	.519**	-						
Raw TLX Performance Low Complexity Yes Redundancy	.171	.068	.114	.048	.475**	.299	-					
Raw TLX Performance High Complexity Yes Redundancy	.040	-.131	.068	.070	.392*	.262	.522**	-				
Raw TLX Mental Effort Low Complexity No Redundancy	.497**	.284	.377*	.412*	.473**	.338	.084	.010	-			
Raw TLX Mental Effort High Complexity No Redundancy	.176	.469**	.350*	.337	.227	.649**	-.041	-.123	.702**	-		
Raw TLX Mental Effort Low Complexity Yes Redundancy	.220	.205	.297	.362*	.344*	.371*	.162	.135	.744**	.686**	-	
Raw TLX Mental Effort High Complexity Yes Redundancy	.180	.169	.252	.393*	.225	.339*	.052	.257	.719**	.630**	.901**	-

**p<0.01
*p<0.05
N=34

.	Total Raw TLX LC-No R	Total Raw TLX HC-No R	Total Raw TLX LC-Yes R	Total Raw TLX HC-Yes R	Raw TLX MD LC-No R	Raw TLX MD HC-No R	Raw TLX MD LC-Yes R	Raw TLX MD HC-Yes R	Raw TLX PD LC-No R	Raw TLX PD HC-No R	Raw TLX PD LC-Yes R	Raw TLX PD HC-Yes R
Raw Frustration Score Low No R	.451**	.321	.359*	.301	.200	.113	.151	.118	-.422*	-.419*	-.402*	-.425*
Raw Frustration Score High No R	.299	.447**	.308	.184	.171	.179	.159	.112	-.460**	-.457**	-.439**	-.464**
Raw Frustration Score Low Yes R	.384*	.368*	.482**	.370*	.182	.186	.191	.144	-.352*	-.350*	-.329	-.357*
Raw Frustration Score High Yes R	.378*	.257	.356*	.426*	.124	.039	.118	.124	-.373*	-.371*	-.351*	-.378*
Percent Correct Low No R	-.453**	-.103	-.066	-.112	-.377*	-.073	.113	.015	-.142	-.156	-.089	-.123
Percent Correct High No R	-.090	-.221	.136	.094	-.299	-.329	.052	-.048	-.062	-.083	.016	-.032
Percent Correct Low Yes R	-.030	-.004	-.143	-.054	.153	.191	.129	.144	-.215	-.203	-.254	-.237
Percent Correct High Yes R	-.143	-.009	-.197	-.249	-.022	.010	.037	-.072	-.202	-.208	-.201	-.199
Mean (SD)	52.47	57.82	40.71	48.85	54.68	59.79	48.26	55.71	6.53	6.41	5.79	6.88
	(13.82)	(13.80)	(12.25)	(13.54)	(19.42)	(18.05)	(18.30)	(19.87)	(18.32)	(18.12)	(16.54)	(19.40)

**p<0.01

*p<0.05

N=34

.	Raw TLX TD LC- No R	Raw TLX TD HC- No R	Raw TLX TD LC- Yes R	Raw TLX TD HC- Yes R	Raw TLX P LC- No R	Raw TLX P HC- No R	Raw TLX P LC- Yes R	Raw TLX P HC- Yes R	Raw TLX ME LC- No R	Raw TLX ME HC- No R	Raw TLX ME LC- Yes R	Raw TLX ME HC- Yes R
Raw Frustration Score Low No R	.119	-.070	.195	.060	.497**	.235	.306	.374*	.195	.111	.111	.001
Raw Frustration Score High No R	-.011	.002	.193	.014	.361*	.319	.207	.188	.111	.261	.078	-.083
Raw Frustration Score Low Yes R	-.032	-.080	.118	.012	.382*	.292	.397*	.417*	.129	.210	.293	.167
Raw Frustration Score High Yes R	.035	-.061	.050	.129	.356*	.183	.298	.471**	.120	.091	.202	.138
Percent Correct Low No R	-.288	-.067	.092	.004	-.513**	-.101	-.176	-.129	-.268	-.001	-.143	-.049
Percent Correct High No R	-.042	-.253	.198	.072	-.147	-.213	.210	.171	-.110	-.231	-.117	.007
Percent Correct Low Yes R	.052	.055	-.053	.108	-.025	-.055	-.337	-.251	.123	.133	.025	.065
Percent Correct High Yes R	.211	.256	.151	.075	-.094	-.082	-.298	-.347*	-.019	.072	-.165	-.144
Mean (SD)	55.65	60.56	46.97	55.74	51.09	58.44	29.09	38.97	55.71	59.76	43.88	51.35
	(17.20)	(18.13)	(16.43)	(18.83)	(15.30)	(16.20)	(12.52)	(14.09)	(17.98)	(18.54)	(17.90)	(18.80)

**p<0.01
*p<0.05
N=34

	Raw TLX F LC-No R	Raw TLX F HC-No R	Raw TLX F LC-Yes R	Raw TLX F HC-Yes R	Percent Correct RB LC-No R	Percent Correct RB HC-No R	Percent Correct RB LC-Yes R	Percent Correct RB HC-Yes R
Raw TLX Frustration Low Complexity No Redundancy								
Raw TLX Frustration High Complexity No Redundancy	.900**							
Raw TLX Frustration Low Complexity Yes Redundancy	.875**	.830**						
Raw TLX Frustration High Complexity Yes Redundancy	.884**	.792**	.895**					
Percent Correct Read-Back Low Complexity No Redundancy	-.171	.001	-.080	-.134				
Percent Correct Read-Back High Complexity No Redundancy	.184	.166	.186	.176	.592**			
Percent Correct Read-Back Low Complexity Yes Redundancy	-.144	-.107	-.123	-.061	.278	-.083		
Percent Correct Read-Back High Complexity Yes Redundancy	-.209	-.073	-.270	-.250	.308	.094	.505**	
Mean (SD)	41.18	43.76	29.68	35.21	72.06	63.32	92.06	83.53
	(25.83)	(25.56)	(19.44)	(22.43)	(11.80)	(13.38)	(5.96)	(8.57)

**p<0.01
*p<0.05
N=34

Cognitive Load and Performance Analyses

Hypotheses 1A and 1B: Effects of Redundancy on Cognitive Load and Performance

Hypothesis 1A proposed that redundancy in the display of verbal information would be associated with lower cognitive load, and no redundancy with higher cognitive load. A series of seven mixed-model ANOVAs was conducted on the cognitive load measures (one for each raw NASA TLX subscale measures and one on the total raw NASA TLX measure). All analyses were performed using SPSS General Linear Model Repeated Measures. An alpha level of .01 was used for all analyses conducted on the six NASA TLX subscale measures and alpha level of .05 was used for the analysis of the total NASA TLX measure.

A significant within-subjects effect was present for redundancy, $F_{Total}(1,30)=67.83$, $p<.005$, *Partial Eta*^{2} =.693, on the total NASA TLX measure. In that, without redundancy, workload was rated higher ($M_{NR\ Total}=56.08$, $SE=2.25$) than with redundancy ($M_{R\ Total}=45.39$, $SE=2.25$). A significant main effect for redundancy was found on five of the six NASA TLX subscales. Workload was rated higher without redundancy and lower with redundancy on all subscales (Table 14).

Table 14 Within-subject effects for redundancy on NASA TLX subscales

NASA TLX Subscale	F_R (1,30)	p	*Partial Eta*^{2}	M_{NR}	SE_{NR}	M_R	SE_R
Mental Demand	9.51	.004	.241	58.13	3.21	52.71	3.40
Temporal Demand	17.40	>.005	.367	57.68	2.97	50.64	3.07
Performance	77.87	>.005	.722	55.88	2.44	34.28	2.06
Mental Effort	21.83	>.005	.421	58.90	3.05	48.89	3.24
Frustration	22.36	>.005	.427	44.11	4.61	33.93	3.66

No significant main effect of redundancy was found on the NASA TLX Physical Demand measure. Figure 16 and Figure 17 illustrate the results of the analyses conducted on the effects of verbal display redundancy on cognitive load. The pattern of scores across all NASA TLX scales showed significantly decreased cognitive load when verbal redundancy was present. The NASA TLX Performance subscale is defined as the subjective assessment of how successful participants think they were in accomplishing the goals of the task and how satisfied were with their performance in accomplishing those goals. The endpoints of this subscale are “Perfect” on the left hand side of the scale, and “Failure” on the right, indicating increase of workload associated with performance from left to right. Therefore, the pattern of NASA TLX Performance scores decreasing when verbal redundancy was present as shown on Figure 10 (left) indicates decreased cognitive load due to the subjective perception of performance as a source of workload and not as a measure of performance per se.

Hypothesis 1B proposed that redundancy in the display of verbal information would be associated with greater performance, and no redundancy with lower performance.. A mixed model ANOVA was conducted on the performance measure (percent correct clearance read-back). The analysis was performed using SPSS General Linear Model Repeated Measures. An alpha level of .05 was used. A significant within-subjects effect was present for redundancy $F_{Ppc}(1,30)=89.72$, $p<.005$, $Partial\ Eta^2=.749$. In that, without redundancy, the percent correctly read back clearance items was lower ($M_{NR\ Ppc}=66.82$, $SE=2.02$) than with redundancy ($M_{R\ Ppc}=87.53$, $SE=1.15$). Figure 18 illustrates the results of the analysis conducted on the effects of redundancy on performance.

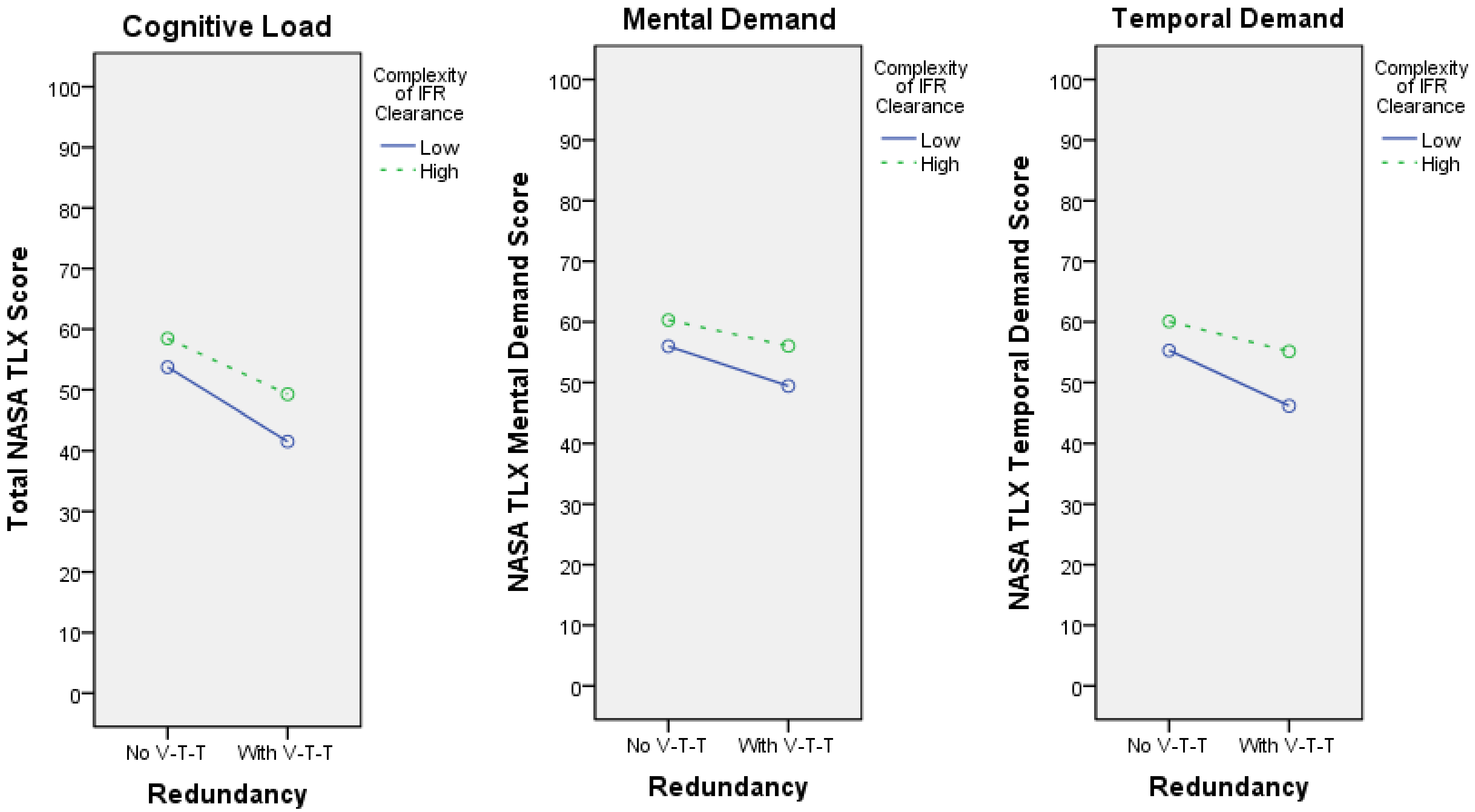


Figure 16 Effects of Redundancy on Cognitive Load (Total Cognitive Load, Mental Demand, and Temporal Demand)

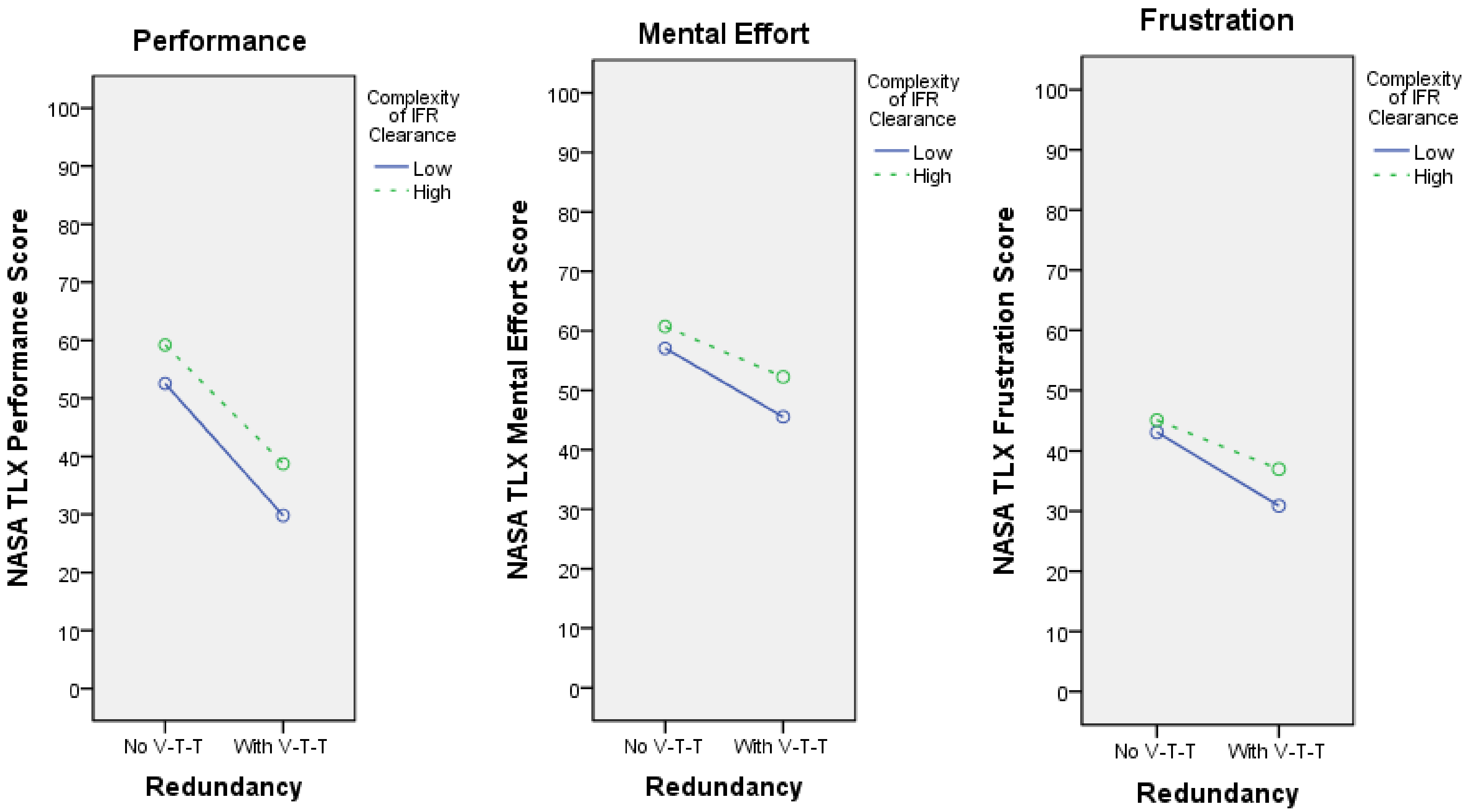


Figure 17 Effects of Redundancy on Cognitive Load (Performance, Mental Effort, and Frustration)

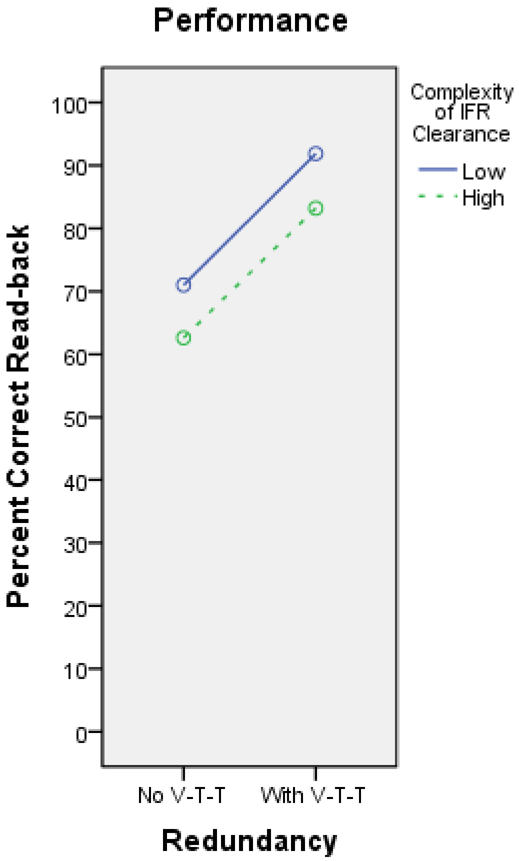


Figure 18 Effects of Redundancy on Performance (Percent Correct Read-back)

Hypotheses 2A and 2B: Effects of Complexity on Cognitive Load and Performance

Hypothesis 2A proposed that higher complexity clearances would be associated with higher cognitive load and lower complexity clearances with lower cognitive load. Similarly to the main effect found for redundancy, the series of seven ANOVAs conducted on the cognitive load measures found a significant within-subjects effect of complexity on the total NASA TLX measures, $F_{C\ Total}(1,30)=28.23$, $p<.005$, *Partial Eta*^{2}=.485. In that, for low complexity clearances, overall workload was rated lower ($M_{LC\ Total}=47.62$, *SE*=2.20) than for high complexity clearances ($M_{HC\ Total}=53.85$, *SE*=2.26). Significant main effects for complexity were also found on five of the six NASA TLX subscales. Workload was rated lower for low complexity clearances and higher for high complexity clearances on all subscales (Table 15).

Table 15 Within-subject effects for complexity on NASA TLX subscale scores

NASA TLX Subscale	F_C (1,30)	p	Partial Eta^2	M_{LC}	SE_{LC}	M_{HC}	SE_{HC}
Mental Demand	23.89	<.005	.453	52.69	3.15	58.16	3.32
Temporal Demand	12.58	.001	.295	50.72	2.87	57.60	3.23
Performance	17.88	<.005	.373	41.18	2.11	48.97	2.12
Mental Effort	13.87	.001	.316	51.31	2.95	56.48	3.12
Frustration	10.81	.003	.265	36.99	3.98	41.04	4.15

No significant main effect of complexity was found on the NASA TLX Physical Demand measure. Table 12 and Table 13 illustrate the results of the analyses conducted on the effects of complexity on cognitive load. The pattern of scores across all NASA TLX scales showed that cognitive load varied with complexity. It was higher when the IFR clearances were more complex and lower when the clearances were less complex. The pattern of NASA TLX Performance scores increasing with complexity as shown on Figure 20 (left) seems counterintuitive. Based on the definition of this NASA TLX subscale, the graph should be interpreted as indicating an increased cognitive load due to the subjective perception of performance as a source of workload, and not as a measure of performance per se.

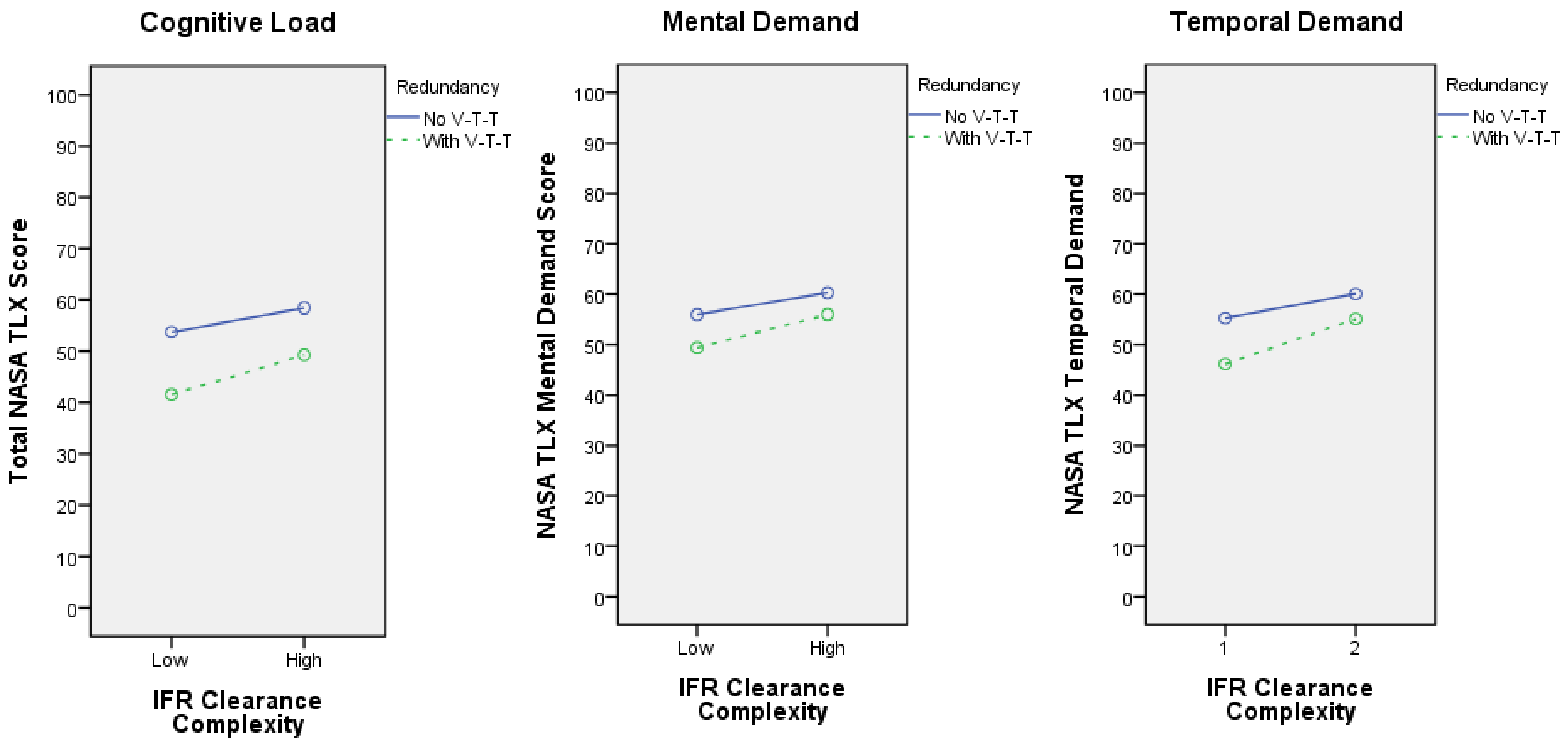


Figure 19 Effects of Complexity on Cognitive Load (Total Cognitive Load, Mental Demand, and Temporal Demand).

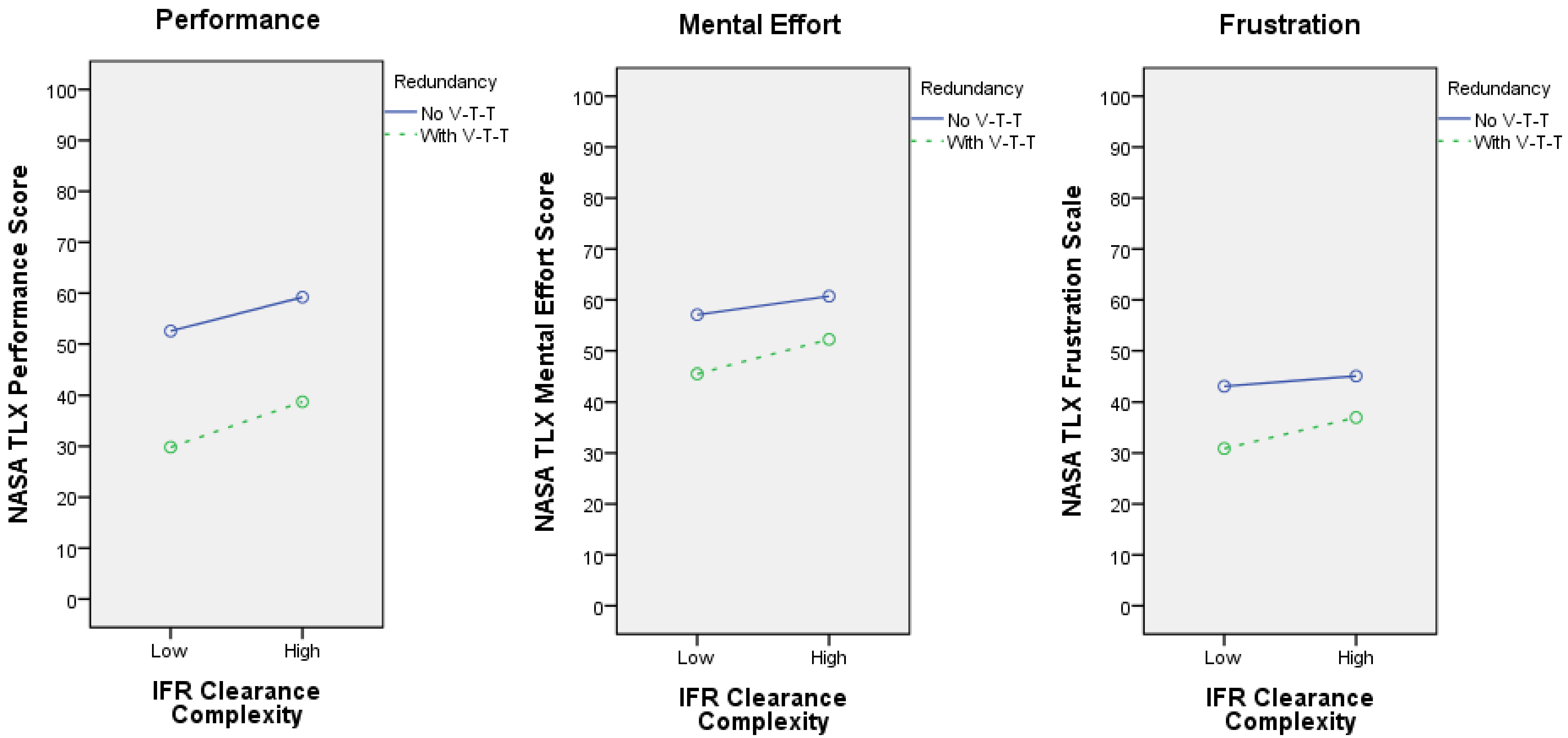


Figure 20 Effects of Complexity on Cognitive Load (Performance, Mental Effort, and Frustration).

Hypothesis 2B proposed that higher complexity clearances would be associated with lower performance and lower complexity clearances with greater performance. A mixed ANOVA was conducted on the performance measure. The analysis was performed using SPSS General Linear Model Repeated Measures. An alpha level of .05 was used. A significant within-subjects effect was present for complexity, $F_{C\ Ppc}(1,30)=44.91$, $p<.005$, *Partial* $Eta^2 =.600$, where $M_{LC\ Ppc}=81.44$, $SE=1.26$, and $M_{HC\ \mathrm{Ppc}}=72.91$, $SE=1.49$. Figure 21 illustrates the results of the analysis conducted on the effects of complexity on performance.

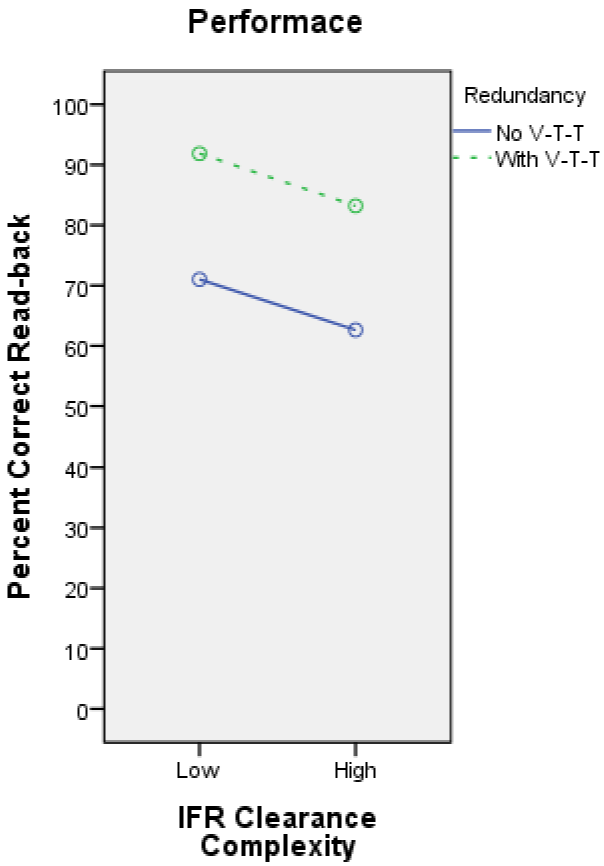


Figure 21 Effects of Complexity on Performance (Percent Correct Read-back).

Hypotheses 3A and 3B: Effects of Level of Expertise on Cognitive Load and Performance

Hypothesis 3A proposed that higher level of expertise would be associated with lower cognitive load and lower level of expertise with higher cognitive load. Furthermore, *Hypothesis 3B* proposed that higher level of expertise was expected to be associated with greater performance and lower level of expertise with lower performance. There were no main effects found in support of these hypotheses when total flight hours were not included in the analyses. Therefore, further analysis was conducted where total flight hours was included as a covariate. The results validated those from the original analysis – no main effects of expertise on cognitive load and performance. These findings were somewhat unexpected and are further discussed in Chapter Six of this dissertation.

Although English as a second language was not included in the literature review section of this dissertation as a variable of interest, a significant portion (38%) of the population of participants who attended the study was with English as a second language (L2). Therefore, all statistical analyses conducted on the data and reported here included L2 as a between-subject variable. The presence of V-T-T exhibited the same beneficial effects for English speaking participants and participants with English as a second language as for experts and novices (Figure 22 and Figure 23). Furthermore, the performance scores (for both novices and experts, as well as native and L2 participants) when the V-T-T transcription was available to the pilot were well into the upper one-quarter of the percent correct scale showing a significantly improved performance when compared to the scores without V-T-T (Figure 22).

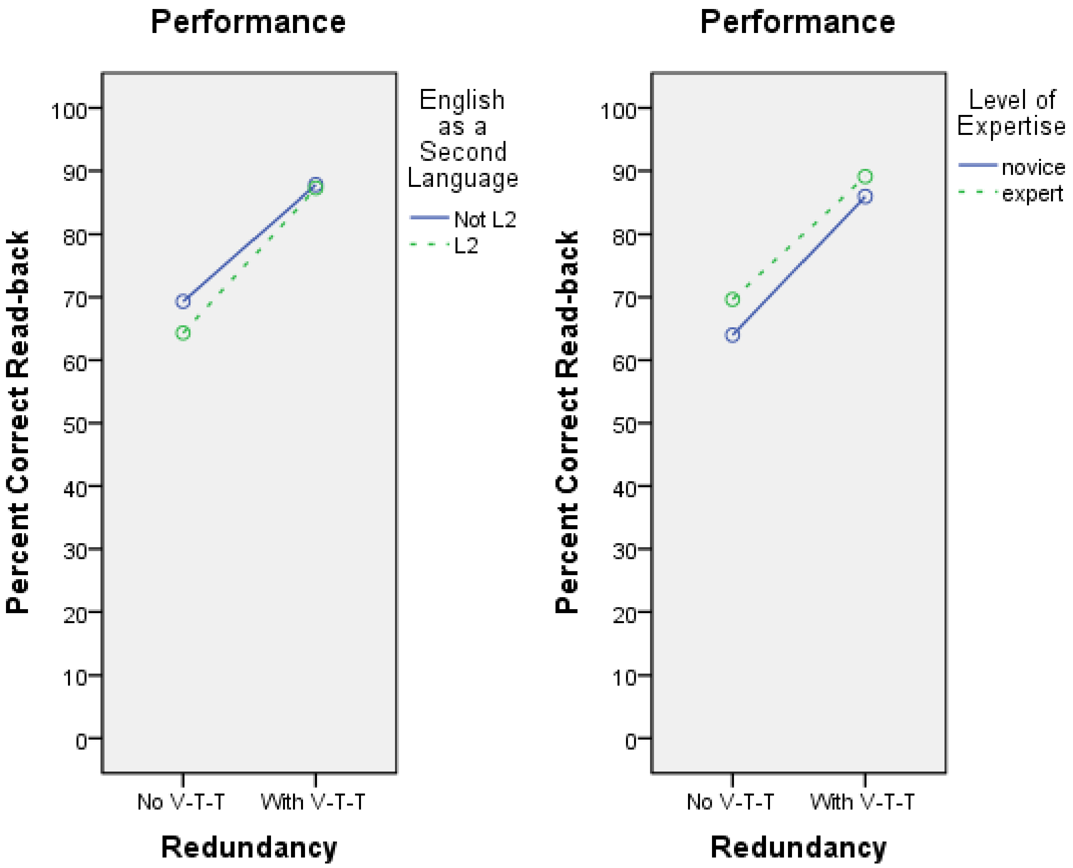


Figure 22 Effects of Redundancy on Performance (Percent Correct Read-back) for Native and English as Second Language Speakers (left) and Level of Expertise (right).

Similarly, the cognitive load scores (for both novices and experts, as well as native and L2 participants), when the V-T-T transcription was available, were in the mid to lower section of the NASA TLX scale, showing a significant decrease in workload for that condition when compared to “No V-T-T” condition (Figure 23).

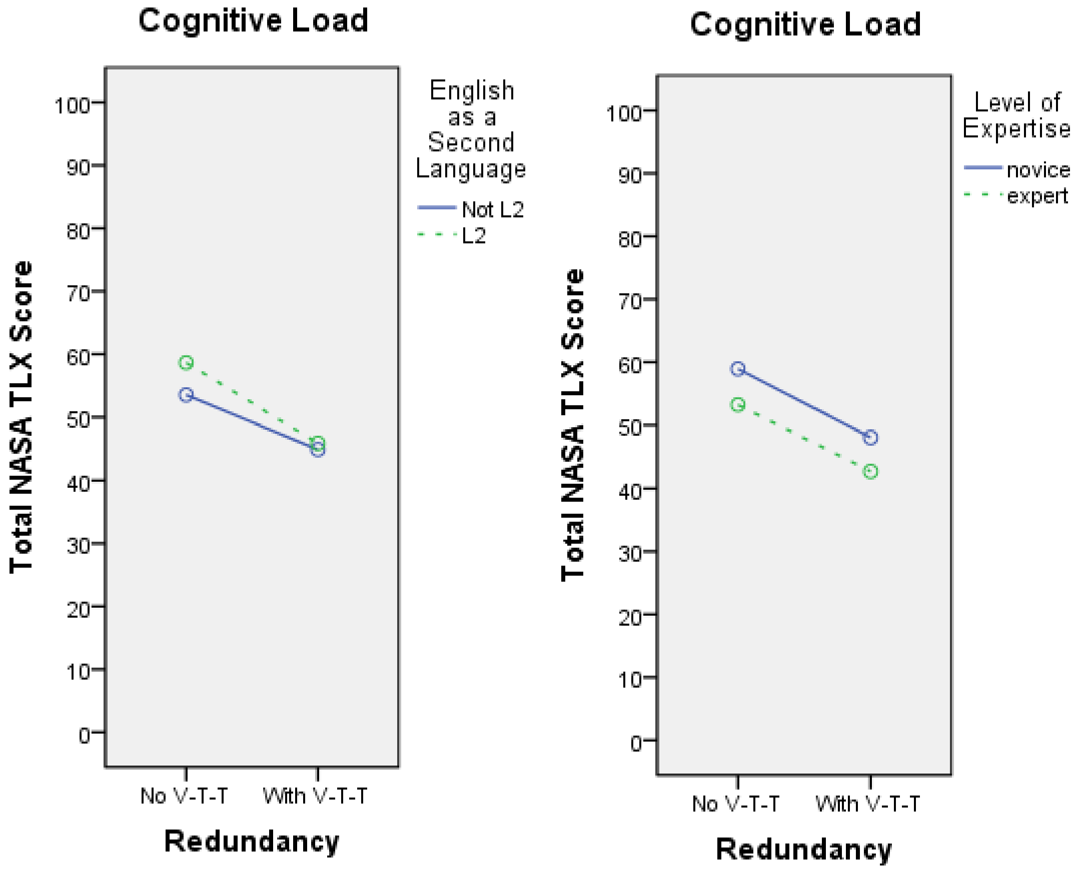


Figure 23 Effects of Redundancy on Cognitive Load (Total NASA TLX Score) for Native and English as Second Language Speakers (left) and Level of Expertise (right).

Hypotheses 4A-F: Interactions between Verbal Display Redundancy, Verbal Input Complexity, and Level of Expertise on Cognitive Load and Performance

Hypothesis 4A proposed that for the no redundancy/high complexity condition, cognitive load would be higher than for no redundancy/low complexity condition regardless of level of expertise. Hypothesis 4B stated that for the redundancy/high complexity condition, cognitive load would be higher than for redundancy/low complexity condition regardless of level of expertise. Furthermore, Hypothesis 4C proposed that in the redundancy/high complexity and redundancy/low complexity conditions, the differences in cognitive load between novices and experts would be significantly reduced compared to the differences between novices and experts in the no redundancy/high complexity and no redundancy/low complexity conditions. Hypothesis 4D proposed that for the no redundancy/high complexity condition, performance would be lower than for no redundancy/low complexity condition, regardless of level of expertise. Hypothesis 4E stated that for the redundancy/high complexity condition, performance would be lower than for redundancy/low complexity condition, regardless of level of expertise.

Furthermore, *Hypothesis 4F* proposed that for the redundancy/high complexity and redundancy/low complexity conditions, the differences in performance between novices and experts would be significantly reduced compared to the differences between novices and experts for the no redundancy/high complexity and no redundancy/low complexity conditions. There were no significant interactions found in support of these hypotheses.

Hypotheses 5: Assessing the Relationship between Manipulated Factors, Cognitive Load, and Performance

Hypothesis 5 stated that the Redundancy X Complexity X Expertise interaction and performance outcome would be mediated by cognitive load. In light of the findings from the statistical analyses conducted for Hypotheses 1 through 4, more specifically, that no significant interactions were found between redundancy, complexity, and expertise on the measures of cognitive load and performance, to test Hypotheses 5, the following analyses per Baron and Kenny (1986) (Figure 24) for the entire group of participants (novices and experts) were conducted to test the moderated mediation model shown on (Figure 10).

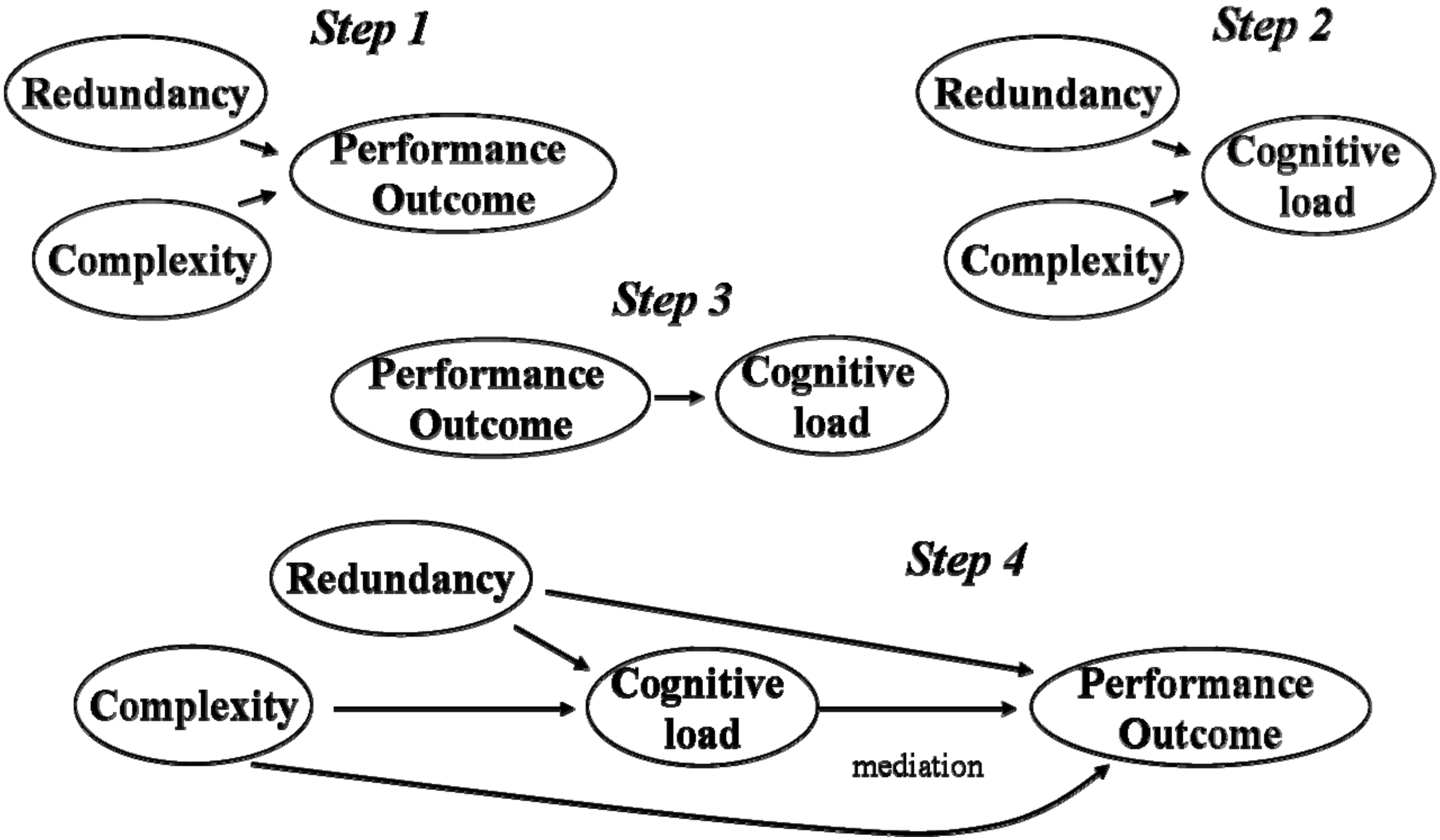


Figure 24 Analyses conducted to test for mediation per Baron and Kenny (1986)

First, the relationship between complexity, redundancy, and their interaction complexity X redundancy, nested within participants, using percent correctly read back clearance items as the measure for performance (dependent variable) was assessed. The analysis showed a significant negative main effect of complexity on performance scores, *F* (1, 99) =28.948, *p*< .005. Inspection of the effect estimates showed a negative effect (*Est*. = -7.97) of complexity on performance. For redundancy, the effect was positive (*Est*. = 20.264) and significant, *F* (1, 99) = 164.013, *p*< .005, but there was no significant interaction between complexity and redundancy. Similar to the results in Step 1, the results of the Step 2 analyses validated the results of the analyses conducted on the performance scores using mixed ANOVAs to test Hypotheses 1 through 4.

Second, the relationship between complexity, redundancy, and their interaction Complexity X Redundancy nested within participants, using NASA TLX as the measure for cognitive load (dependent variable) was assessed using HLM (SPSS Mixed Models – Linear). The analysis showed a significant positive association of complexity on cognitive load (NASA TLX) scores, *F* (1, 99) =31.01, *p*< .005. Inspection of the parameter estimates showed a positive relationship of complexity. Under high complexity, the cognitive load (NASA TLX) parameter was higher (by an estimated 8.14 points) than under low complexity. For redundancy, the analysis showed a significant negative association, *F* (1, 99) = 73.17, *p*< .005, but no significant interaction between complexity and redundancy was found. The results of these analyses validated the results of the analyses conducted on the cognitive load scores using mixed model ANOVAs to test Hypotheses 1 through 4.

Third, the relationship between cognitive load and performance was tested using a hierarchical model with cognitive load (NASA TLX) as a fixed-effects covariate. The analysis showed a significant negative association of cognitive load and performance scores, *F* (1, 133.889) =120.441, *p*< .005. Inspection of the parameter estimates indicated a negative effect (*Est.* = -.942) of cognitive load on performance.

Forth, because the hypothesized moderated mediation model was not supported by the results from the Hypotheses 3 through 4 testing (i.e., no significant interaction between complexity and redundancy was found) a revised mediation model of relationships between variables used in the study was assessed. The results are shown in Table 16.

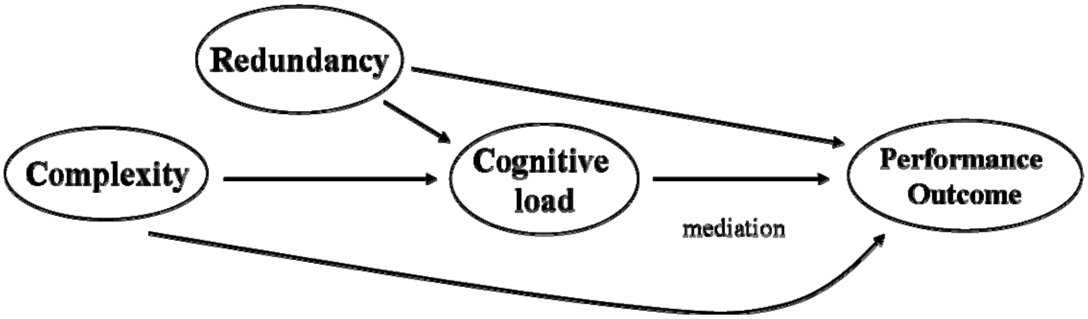


Figure 25 Revised model

Table 16 Results from analysis conducted in Step 4

	F (df)	*p*	*Estimate*
Complexity	15.021 (1, 106.649)	<.005	-4.77
Redundancy	97.49 (1, 120.194)	<.005	16.74
Cognitive Load	23.90 (1,90.69)	<.005	-.392

Dependent variable: Performance (Percent correct read back)

Although cognitive load remained as a significant negative predictor of performance, both complexity and redundancy also had unique significant effects on performance suggesting only partial rather, than full mediation.

Usability of V-T-T Transcription Analyses

A usability survey (Appendix E) was conducted to gauge participants' perception of function (i.e., if the V-T-T transcription functioned as intended), format (e.g., shorthand abbreviations and font size), the ease of use, ease of interpretation of the V-T-T transcription widget used in the study. A five-point scale was employed, where 1=”Very Poor”, 2=”Poor”, 3=”Acceptable”, 4=”Good”, and 5=”Very Good” for function and format; and 1=”Very Difficult”, 2=”Difficult”, 3=”Neutral”, 4=”Easy”, and 5=”Very Easy” for ease of use and ease of interpretation. The results of this survey are shown on Figure 26 below.

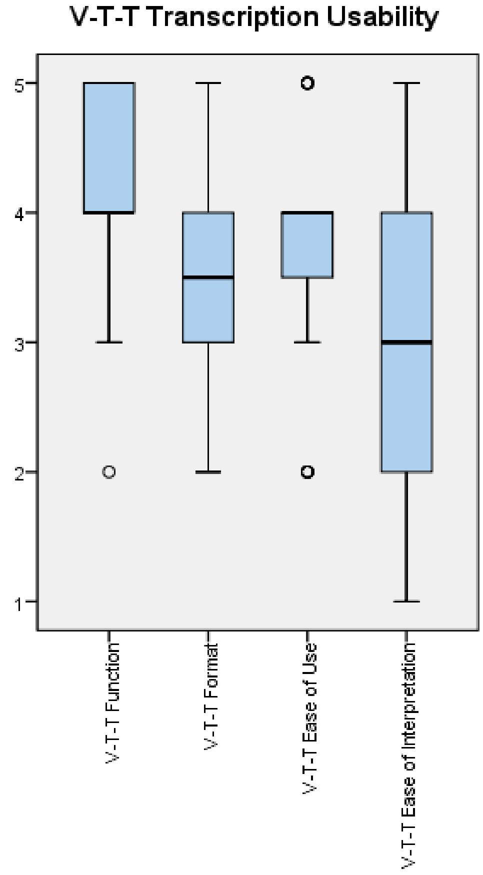


Figure 26 Results of the V-T-T Usability Survey

While the responses were mostly positive with regards to the V-T-T transcription's function, format, and ease of use as presented during the experiment, the pattern of responses reflecting on the ease of interpretation was not as unambiguous. For example, some of the abbreviations/symbols contained in the shorthand method utilized for the V-T-T transcription were noted by the participants as having a very ambiguous meaning. Almost all of the participants had difficulty interpreting the "less than" (<) symbol as "AFTER" (e.g., "< DP" = "after departure") and the "greater than" (>) as "BEFORE (reaching, or passing)". At the same time, participants had no issues interpreting symbols like “up arrow” (↑), “down arrow” (↓), and “right arrow” (→) as “climb”, “descend”, and “cruise”, respectively. Such results may be explained with the lack of existing standardized shorthand for capturing clearances, as well as the lack of mandatory shorthand training as part of pilot training and education at large.

CHAPTER SIX: DISCUSSION

The goals of this dissertation research were (a) to examine the utility of verbal display redundancy in managing pilot cognitive load during controller-pilot voice communications, (b) to test the validity of applying a theoretical framework, which stemmed from the domain of learning and instruction, and (c) to ultimately extend that framework, into operational domains (e.g., a flight deck). A controller-pilot voice communications task (e.g., reading back an IFR clearance) was adapted to test these goals. The importance of the study as a stepping stone for addressing the utility of verbal redundancy as a potential solution for managing pilot cognitive load and the larger implications of such solution for reducing errors of miscommunications, as well as, the potential for expanding the theoretical framework into the domain of flight operations (e.g., controller-pilot voice communications) are discussed in this chapter.

Although English as a second language (L2) was not considered as a variable of interest for this dissertation, due to naturally occurring diversity of native languages in the pilot population at large (which was reflected in the sample population used in the experiment), the same response pattern was observed for native English speakers and L2 participants. The practical implication of this finding will be discussed later in this chapter.

The chapter is organized as follows. First is a summary of the results by hypothesis. A discussion on the theoretical implications of the research is next, which is followed by a discussion on the practical implications of the research. Study limitations and future research questions are addressed, followed by a closing with concluding remarks about the research.

Hypothesis Discussion

Hypothesis 1: Effects of Redundancy on Cognitive Load and Performance

The focus of Hypotheses 1 was to investigate the effects of redundancy on cognitive load and performance. In the presence of redundant voice-to-text transcription, pilot cognitive load was significantly lower, and performance was significantly better. The difference in scores between no redundancy and redundancy conditions was in the range of 7-11 points on the total NASA TLX workload score and similar for the scores on most of the NASA TLX subscales. These scores reside in the middle of the workload scales indicating a change in workload from moderately high in the no redundancy condition to moderately low in the redundancy condition. However on the NASA TLX Performance subscale the difference in scores between those two conditions was more than 20 points, which is more than 20% reduction of workload on that subscale (Figure 17). Similarly, the difference in percent correctly read back clearance items between no redundancy and redundancy conditions was also more than 20 points, which accounts for more than 20% improvement in performance. These findings, and particularly the very similar pattern of scores between the no redundancy and redundancy conditions on the NASA TLX Performance subscale and the scores on objective measure of performance suggest that reading back a clearance correctly is perceived by the pilot community at large as an essential for the safety of flight skill, and therefore they are more likely to have a very accurate self assessment of how well they did on that task.

Hypothesis 2: Effects of Complexity on Cognitive Load and Performance

Hypothesis 2 addressed the effects of IFR clearance complexity on cognitive load and performance. The results are not difficult to interpret – in the high complexity condition, participants reported higher cognitive load than in the low complexity condition, and their performance scores were lower in the high complexity and higher in the low complexity condition. The difference in workload scores between no redundancy and redundancy conditions was consistently less than 10 points on the total NASA TLX workload score and the scores on all subscales. In the realm of IFR clearances, the complexity of even the simplest of clearances is still pretty substantial as compare to visual flight rules (VFR) clearances, for example. However the results of comparing two levels of complexity from two different domains of clearances in the context of researching the utility of verbal redundancy would have been misleading due to the different levels of element interactivity within each of these domains. More importantly, most of the flying in the National Airspace is conducted under IFR, which makes the selection of clearances for this dissertation research operationally very relevant and with solid ecological validity.

Hypothesis 3: Effects of Expertise on Cognitive Load and Complexity

Hypothesis 3 focused on level of expertise and its effects on cognitive load and performance. It was hypothesized that V-T-T transcription of the ATC clearances would benefit mostly less experienced pilots, and that the benefits for more experienced pilots would be limited. The results did not support the latter. Rather, the pattern of responses reflected consistently lower cognitive load and improved performance for both novices and experts in the verbal display redundancy condition. This particular result can be attributed to the research paradigms employed for this dissertation, i.e. dual-task (primary and secondary), and Brown-Peterson (Brown, 1958; Peterson & Peterson, 1959) within the primary task. The intent behind utilizing Brown-Peterson paradigm was to prevent rehearsal and ultimately narrow down the results to only reflect the effects of the manipulated variables on cognitive load and performance. When combined with the choice of range between low and high complexity IFR clearances, the deployment of Brown-Peterson paradigm “limited” the effects of expertise to reside only within the “boundaries” of working memory as identified by the Baddeley’s model (Baddeley, 1981; 1986; 1992; 1996; 1998; Baddeley & Hitch, 1994) and therefore pose no significant impact on cognitive load or performance.

Hypotheses 4: Interactions between Redundancy, Complexity, and Expertise

Hypothesis 4 proposed that there would be significant interactions between redundancy, complexity, and expertise. No significant interactions were found in support of this hypothesis. The hypothesized differential impact of complexity and redundancy on level of expertise was not supported. The availability of V-T-T was equally beneficial in improving performance and reducing cognitive load for novices and experts. Within the realm of IFR clearances the first level of complexity (low) was difficult enough so that redundancy aided both novices and experts. The second level (high) of complexity was even higher and the participants across levels of expertise were aided by the presence of V-T-T transcription, as well. Such result may be attributed to the large main effects of redundancy and complexity on cognitive load and performance when considered in the context of the participants’ selection criteria for novices and experts. While a different set of selection criteria (e.g., lower upper end of flight hours for novices and higher lower end of flight hours for experts) might have yielded significant interactions, such set of criteria would have not been sensible nonetheless. The utility of any design solution should be researched such that it accounts for the majority of its target user population and without impractical exclusions. In addition, based on experimenter’s observations and verbal feedback from some of the participants (>7000 total flight hours) in the expert pilot group, the utility of V-T-T transcription for these populations may be geared more towards helping them verify they had captured the clearance correctly rather than relying on it as a primary means for recording it.

Hypothesis 5: Moderated Mediation between Complexity, Redundancy, Cognitive Load and Performance

A moderated mediation was hypothesized to exist between complexity, redundancy, expertise, cognitive load, and performance. In particular, it was hypothesized that the Redundancy X Complexity X Expertise interaction and performance outcome would be mediated by cognitive load. Following Baron and Kenny (1986), support for the moderation portion of the hypothesis was predicated on a significant interaction existing between complexity, redundancy, and expertise. No significant interaction was found between these variables.

Since the mediation portion of the impact of complexity and redundancy on performance by cognitive load, was of continued interest, the originally proposed moderated mediation model was revised. Instead, I tested whether cognitive load mediated the impact of the main effect of complexity and redundancy on performance by conducting a sequence of hierarchical linear models in line with Baron and Kenny's (1986) approach. The tests were conducted on the full sample of participants. The results from testing the revised model negated full mediation and instead suggested only partial mediation. Thus, they were inconclusive regarding the mediating role of cognitive load in the relationship between the predictor variables (complexity and redundancy) and the outcome variable (performance). The theoretical implications of such findings are discussed next.

Theoretical Implications

The mediating role of cognitive load in the relationship between presentation format, complexity of instructional material, and learning outcomes has been well documented in the instructional research literature (Paas & van Merrienboer, 1994). Theoretically, it was important to verify whether cognitive load plays the same or a similar role in the relationship between these variables in the context of an operational environment such as a modern flight deck, which made it particularly important because the theoretical models employed in this dissertation originated from instructional research domain. The following assumptions about the common attributes between a multimedia learning environment and a modern flight deck were employed. First, from a type of environment stand point, multimedia learning and a modern flight deck are two environments very rich in multimedia. Second, from a task performance standpoint, the execution of complex cognitive tasks such as learning or piloting requires real-time, active processing of new information within the working memory. Third, from a presentation of information standpoint, two main factors influence the ease with which instructions are understood in either of these environments: (a) presentation format and (b) intrinsic complexity.

This dissertation manipulated redundancy (i.e., presentation format), complexity, and level of expertise. A moderated mediation was hypothesized to exist between these variables, cognitive load, and performance outcome. More specifically that the Redundancy X Complexity X Expertise interaction and performance outcome would be mediated by cognitive load. According to Muller et al. (2005), in a moderated mediation, there is an overall treatment effect, where the magnitude of this effect does not depend on the moderator, and only the strength of the mediating process depends on the moderator. Furthermore, moderation is predicated on a significant interaction between the predictor variable (complexity) and the moderator variables (redundancy and expertise) (Baron & Kenny, 1986). No such interaction was found. However the mediating role of cognitive load in the relationship between these variables, and performance was explored further. The results of the tests of the revised model indicated that although cognitive load remained as a significant negative predictor of performance both complexity and redundancy also had unique significant effects on performance, suggesting only partial rather than full mediation. There are several potential explanations for such result.

First, according to Judd and Kenny (2001) a variable may serve as a mediator of the treatment effect if it is causally prior to the outcome variable. Therefore, the conclusions from a mediation analysis are only valid if, in addition to all of the standard assumptions of the general linear model, the causal assumption is met. When the initial variables are manipulated variables, neither the mediator, nor the outcome can cause them. However, precisely because both the mediator and the outcome variables are not manipulated variables, they may cause each other. Although the direction of causation between the mediator and outcome variables cannot be determined by statistical analyses, reverse causation may be ruled out theoretically and by the use of certain research design methods, which can help determine whether the mediator may be caused by the outcome variable. For example, if at all possible the mediator should be measured temporally before the outcome variable. The theoretical background for this dissertation was very robust in terms of the mediating role of cognitive load in the relationship between design and performance in a learning environment. However, the data collected on the measures of cognitive load was retrospective. That is, the subjective assessment of cognitive load (NASA TLX) was conducted after the task was complete; while the data on the measures of performance on the primary task were collected while the task was in progress. Although the NASA TLX was conducted immediately after each of the 20 clearances was read back, it was nonetheless a retrospective measure of cognitive load. A secondary task in this dissertation research was introduced for two reasons, (a) to maintain a constant memory load for the visual-spatial component of working memory; and (b) to serve as an objective measure of cognitive load on the primary task. However, the data collected on secondary task performance, reflected only the overall percent correctly recorded instances when the annunciator light was ON during the test sequence. Such data were considered of a very limited value as a measure of cognitive load due to its low resolution in terms of objectively measuring performance on secondary tasks for each of the four conditions (Table 7), and was therefore abandoned. More specifically, the primary and secondary task sets of stimuli, as well as, the time between each stimulus in these two sets varied in a random manner, therefore capturing, and more importantly linking, these four sets of time stamps, was going to significantly increase the complexity of the simulation software, and due to the very limited resources available for the study, it was deemed impractical for the purposes of this dissertation. Nonetheless, a more accurate secondary task performance as a measure of cognitive load on the primary task, would have afforded a very valuable insight to the relationship between the variables included in the revised model.

Practical Implications

One of the two main goals of this research was to empirically examine the utility of verbal display redundancy in managing pilot's cognitive load during controller-pilot voice communications. The notion that a redundant display of the ATC controller's message in a form of voice-to-text transcription would serve as a memory aid to the pilot, a way to verify a clearance has been captured correctly without having to make a "Say again" call, and has the potential to ultimately improve the margin of safety by reducing the propensity for human error was unequivocally supported by the results from both the utility (cognitive load and performance) measures, and the usability survey administered at the end of each session after the test sequence was completed. The amounts of in reduction of cognitive load and improvement of performance when verbal display redundancy was provided were not trivial at all, but instead in the range of about 20%. They demonstrated the tremendous potential such design solution might have in reducing miscommunications between controllers and pilots in light of the remaining role of voice radio in air traffic management and after the full implementation of CPDLC in the national and international airspace.

Interestingly, almost all the pilots who participated in the study asked the question why this "simple and obvious" solution has not been implemented already. The answer to this question, however, is not simple and obvious. First, it is difficult not to think about the display of V-T-T transcription examined in this study, and a CPDLC display, as being one and the same, unless a very close familiarity with the differences in intended function between the two exists. Second, unfortunately although there has been a tremendous progress made in the last couple of years in improving the accuracy of voice recognition technology required for such application, its bad reputation still persists. More, if a V-T-T transcription were to be made autonomous (no external ground-based infrastructure required) and in order to further improve the accuracy of the voice recognition engine above 95%, many experts in the field recommend the development of an acoustic model for each aircraft's flight deck where implementation of a V-T-T transcription display is desired. Because of the significant investment required, there is reluctance in the industry to go forward with such project. The results of this study however present solid empirical evidence (as opposed to a collection of opinions) about the goodness of providing pilots with such device.

Although there was a very significant 20% improvement in read-back accuracy and all the scores in the redundancy condition were in the upper 20% of the accuracy scale, they also showed that the presence of V-T-T did not produce perfect, or near perfect, accuracy in clearance read-back. These findings will be discussed next in the context of the limitations and future research.

Limitations and Future Research

A substantial effort was made to minimize study limitations through design however there are a few limitations worth discussing. Although these limitations were determined not to be severe enough to confound the results, they should still be taken into account when considering the generalizability of the study. Where applicable, suggestions for future research to address these limitations are made.

Voice Recognition Technology Accuracy

Since the focus of this dissertation was on a subset of the human factors aspects associated with the utility of verbal display redundancy, one of the major assumptions made for the research was that the accuracy of the V-T-T transcription was 100%. Future research topics related to voice recognition technology accuracy in this context may include:

- The impact of different levels of voice recognition accuracy on pilot workload and performance
- Trust (in technology) and individual and/or team performance
- Human error analysis.

Background Chatter

Another limitation of the research was the content and duration of the prerecorded background chatter (also known as "party line") used in the X-Plane® simulation. The default chatter, which comes with the home edition of the software is repetitive, and although relative long in duration (without repetition of the same controller-pilot exchanges), it was nevertheless noted by two of the pilots from the expert group as contributing to a slightly elevated level of frustration. All participants were briefed before the start of the test that these exchanges were prerecorded in a different (from the Dallas-Fort Worth area) part of the country, and only serve as an emulation of real world com radio chatter. None of the pilots noted this as an issue during the test; most likely because the entire experiment was conducted in automatic flight and they were briefed to pay no attention to the content of the "party line" only listen for their call sign.

Clearances play-back

The next limitation of the study was that the participants were not allowed to ask for a clearance play back, i.e., no "Say again" was permitted. While aware of the potential impact of such experimental design decision on the ecological validity of the study, the trade-off here was more experimental control in order to narrow down the differences in cognitive load and performance due to potentially only the presence or absence of V-T-T transcription.

V-T-T Format

Based on a feedback from many instructor pilots who participated in the study, no standard shorthand exists that they were aware of, and no shorthand methods are taught in pilot schools as part of the curriculum. All participants, without exception, reported that the shorthand they use when flying is the shorthand they were first introduced to, by their instructor, and which they have further developed as their own way of jotting down clearances. However, in order to develop the V-T-T transcription display for this study, a shorthand method which utilized common and industry accepted abbreviations and symbols was needed. The method, which met these requirements, was the one introduced in the Jeppesen *ATC Clears* instructional CD. Over 70% of the listed abbreviations and symbols were identified by the pilot participants in the study as very commonly used. Although participants were given plenty of time to review and become fully familiar with the list during the training session conducted before the actual tests, and a laminated copy of these abbreviations and symbols was provided for the duration of the tests; some of them stated that it was still somewhat different from their individual shorthand methods. Those participants described this as causing some frustration and they wished they were trained on a single standard method of shorthand, or had been familiar with the one used in the study, prior to attending. As noted in the previous section the presence of V-T-T did not produce perfect or near perfect accuracy in reading back clearances. This outcome may warrant future investigation of the effects of “training” vs. “no training” conditions on pilot workload and performance. Future research topics related to the format of the V-T-T display may also include the use symbols vs. abbreviations vs. full transcription (where all identifiers of airports, navigation aids, etc., are presented with their full name, e.g., Dallas-Fort Worth instead of DFW, or presented as both the identifier and full name). In addition, the ability for the pilots to control the display of V-T-T transcription in terms of format, duration, timing, etc., should be further researched.

English as a Second Language

Although it was included in all of the analyses as a between-subject variable, English as a second language (L2) was not a variable of interest in this study. The participants, both L2 and those with English as a native language, unanimously agreed that V-T-T transcription was very helpful for all participants however those who benefited the most were the L2 pilots (both novices and experts). Further research where L2 is a variable of interest should highlight the design features of V-T-T transcription display (e.g., format) that have the most potential for improving performance for L2 participants.

Built-in Time Delay

Another limitation of the study was the built-in delay between the end of the distractor task and the actual showing of the V-T-T transcription (if available) on its dedicated widget within the X-Plane® software. This constant delay was set to 3 sec to simulate the necessary system processing time. However, for a couple of participants this built in delay was perceived as too long. A potential future system implementation of a verbal display of the kind researched here should consider a system requirement for a significantly shorter delay between the end of the message delivered by voice and the display of the V-T-T transcription.

Number of Clearances

Finally, it is worth noting that although the task was intended to emulate the real world task of reading back an IFR clearance as close as possible, the use of 20 arrival and departure clearances within an hour long “flight” using a desk top simulation was far from realistic. Knowing in advance that this may potentially be noted by the participants as an issue, and considering the logistics involved in the study it was determined that this limitation was an acceptable task realism compromise. In addition, because all participants were briefed about this trade-off, none of them expressed any concern regarding the about the task realism.

Secondary Task

A dual-task methodology was utilized in this dissertation study, where the primary task was to listen and read back ATC clearances (verbal) and the secondary task was to monitor a cockpit indicator light (visual). For the secondary task, the participants were asked to continuously monitor a green annunciator light on the upper instrument panel of the cockpit in the X-Plane® simulation and write down the times the light was ON by recording the time in minutes and seconds from the digital clock provided. Two approaches may be applied in a dual-task paradigm. One approach is to add a secondary task with the only intent of introducing memory load. Another approach is to use secondary task performance as a measure of the cognitive load induced by the primary task. Due to the limited resources available for this study, a more accurate way of capturing real time objective performance data on the secondary task was not feasible. Future research where performance on a secondary task is utilized as an objective measure of cognitive load on a primary task should be conducted to further illuminate the role of cognitive load in the relationship between design and performance.

Conclusions

The goals of this study were (a) to investigate the utility of a voice-to-text transcription (V-T-T) of ATC clearances in managing pilot's cognitive load during controller-pilot communications within the context of a modern flight deck environment, and (b) to validate whether a model of variable relationships which is generated in the domain of learning and instruction would "transfer", and to what extend, to an operational domain. First, within the theoretical framework built for this dissertation, all the pertaining factors were analyzed. Second, by using the process of synthesis, and based on guidelines generated from that theoretical framework, a redundant verbal display of ATC clearances (i.e., a V-T-T) was constructed. Third, the synthesized device was empirically examined.

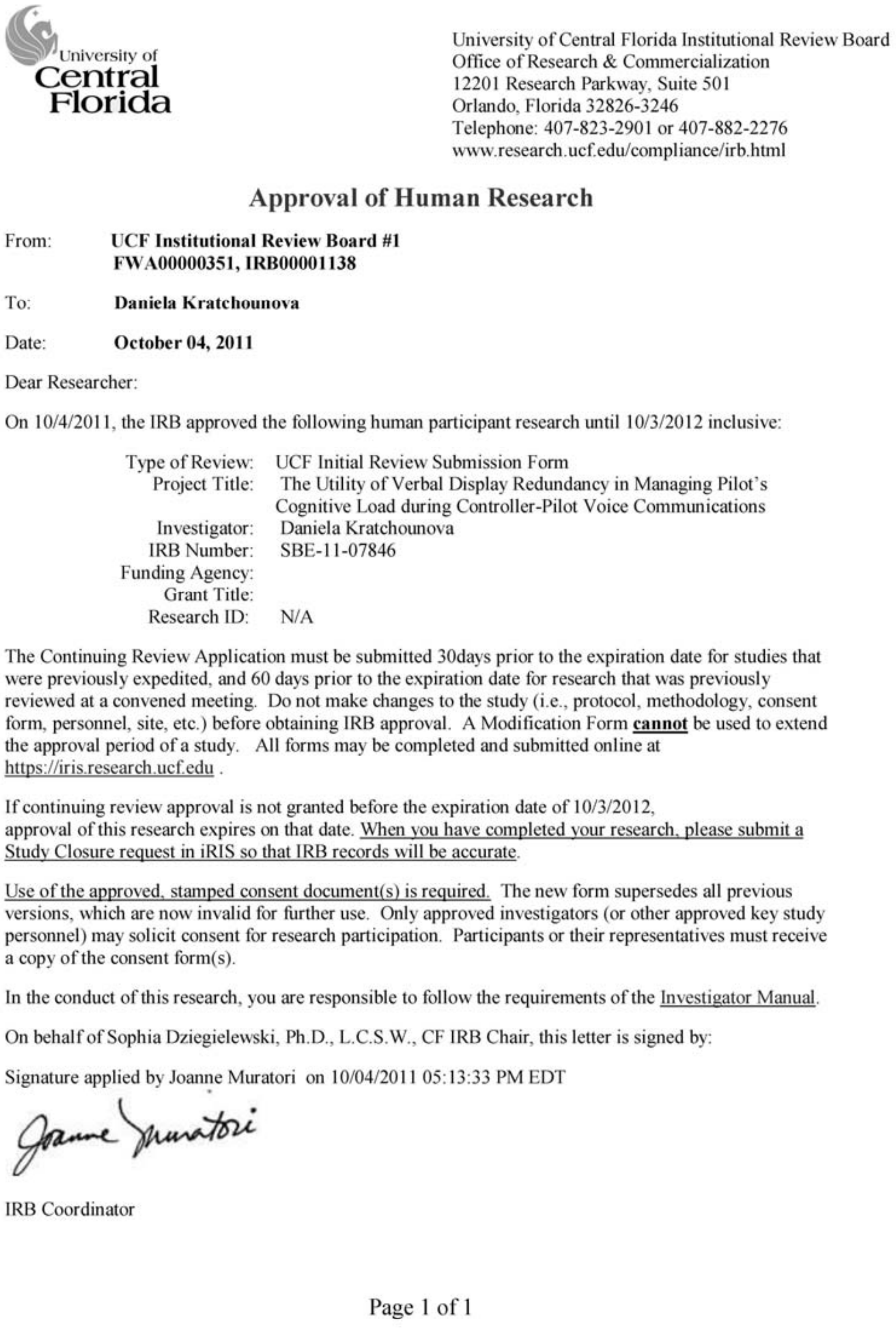
The results showed that the amounts of reduction of cognitive load and improvement of performance, when verbal display redundancy was provided, were in the range of about 20%. These results indicated that V-T-T is a device which has a tremendous potential to serve as (a) a pilot memory aid, (b) a way to verify a clearance has been captured correctly without having to make a "Say again" call, and (c) to ultimately improve the margin of safety by reducing the propensity for human error for the majority of pilot populations including those with English as a second language.

Fourth, the results from the validation of theoretical models "transfer" showed that although cognitive load remained as a significant predictor of performance, both complexity and redundancy also had unique significant effects on performance. These results indicated that the relationship between these variables was not as "clear-cut" in the operational domain investigated here as the models from the domain of learning and instruction suggested.

Furthermore, such results only reinforce the notion that the relationship between cognitive load and performance is multifaceted and very complex in operational environments. For instance, Hancock et al. (1995) caution researchers who use workload measures against the nonlinearity of human response and specifically that there are occasions when a subjective response indicates that the task is becoming more demanding, while at the same time performance is improving and vise versa. The authors point out that people in general, and pilots in particular, use previous experience and future expectations to assess current events. This implies that current events are assessed based on memory rather than on instantaneous change in conditions. In modern flight decks, because the monitoring tasks are rapidly becoming predominant, the finding that a more direct association between cognitive load and performance exists for monitoring tasks is very encouraging (Hancock et al., 1995).

Until further research is conducted, (a) to investigate how changes in the operational task settings via adding additional coding (e.g., permanent record of clearances which can serve as both a memory aid and a way to verify a clearance is captured correctly) affect performance through mechanisms other than cognitive load; and (b) unless the theoretical models are modified to reflect how changes in the input variables impact the outcome in a variety of ways; a degree of prudence should be exercised when the results from the model "transfer" validation are applied to operational environments similar to the one investigated in this dissertation research.

APPENDIX A: IRB APPROVAL LETTER



University of Central Florida Institutional Review Board
Office of Research & Commercialization
12201 Research Parkway, Suite 501
Orlando, Florida 32826-3246
Telephone: 407-823-2901 or 407-882-2276
www.research.ucf.edu/compliance/irb.html

Approval of Human Research

From: **UCF Institutional Review Board #1**
FWA00000351, IRB00001138

To: **Daniela Kratchounova**

Date: **October 04, 2011**

Dear Researcher:

On 10/4/2011, the IRB approved the following human participant research until 10/3/2012 inclusive:

Type of Review:	UCF Initial Review Submission Form
Project Title:	The Utility of Verbal Display Redundancy in Managing Pilot's Cognitive Load during Controller-Pilot Voice Communications
Investigator:	Daniela Kratchounova
IRB Number:	SBE-11-07846
Funding Agency:	
Grant Title:	
Research ID:	N/A

The Continuing Review Application must be submitted 30days prior to the expiration date for studies that were previously expedited, and 60 days prior to the expiration date for research that was previously reviewed at a convened meeting. Do not make changes to the study (i.e., protocol, methodology, consent form, personnel, site, etc.) before obtaining IRB approval. A Modification Form **cannot** be used to extend the approval period of a study. All forms may be completed and submitted online at https://iris.research.ucf.edu .

If continuing review approval is not granted before the expiration date of 10/3/2012, approval of this research expires on that date. When you have completed your research, please submit a Study Closure request in iRIS so that IRB records will be accurate.

Use of the approved, stamped consent document(s) is required. The new form supersedes all previous versions, which are now invalid for further use. Only approved investigators (or other approved key study personnel) may solicit consent for research participation. Participants or their representatives must receive a copy of the consent form(s).

In the conduct of this research, you are responsible to follow the requirements of the Investigator Manual.

On behalf of Sophia Dziegielewski, Ph.D., L.C.S.W., CF IRB Chair, this letter is signed by:

Signature applied by Joanne Muratori on 10/04/2011 05:13:33 PM EDT

Joanne Muratori

IRB Coordinator

Page 1 of 1

APPENDIX B: APPROVED INFORMED CONSENT FORM

The Utility of Verbal Display Redundancy in Managing Pilot's Cognitive Load during Controller-Pilot Voice Communications

Informed Consent

Principal Investigator(s): *Daniela Kratchounova, Ph.D.Candidate*

Faculty Supervisors: *Florian Jentsch, PhD*
Peter Hancock, PhD
Investigational Site: Embry-Riddle Aeronautical University (ERAU)
Human Factors and Systems Department
Daytona Beach, Florida

Introduction: Researchers at the University of Central Florida (UCF) study many topics. To do this we need the help of people who agree to take part in a research study. You are being invited to take part in a research study which will include about 34 people at ERAU. You have been asked to take part in this research study because you are a pilot who either:

- is a holder of at least a Private Pilot License;
- has flown in the last 60 days; and
- has between 100 and 250 total flight hours; ***OR***
- is a holder of at least an Instrument Pilot License;
- has flown in the last 60 days; and
- has at least 500 total flight hours.

You must be 18 years of age or older to be included in the research study.

The person doing this research is Daniela Kratchounova of University of Central Florida, Department of Psychology. Because the researcher is a graduate student, she is being guided by Dr. Florian Jentsch and Dr. Peter Hancock, her UCF faculty supervisors in the Department of Psychology.

What you should know about a research study:

- Someone will explain this research study to you.
- A research study is something you volunteer for.
- Whether or not you take part is up to you.
- You should take part in this study only because you want to.
- You can choose not to take part in the research study.
- You can agree to take part now and later change your mind.

University of Central Florida IRB
IRB NUMBER: SBE-11-07846
IRB APPROVAL DATE: 10/4/2011
IRB EXPIRATION DATE: 10/3/2012

- Whatever you decide it will not be held against you.
- Feel free to ask all the questions you want before you decide.

Purpose of the research study: The purpose of this study is to empirically examine verbal display redundancy as a solution for limiting pilot's cognitive load to only the essential amount of processing resources required for understanding the ATC clearance, and correctly executing the required action, in the larger context of a modern flight deck.

What you will be asked to do in the study: Prior to conducting any portion of the experiment, the Principal Investigator will discuss the informed consent process with you. You will then be briefed on the purpose of the experiment and will be asked to fill out the demographics form. The experiment will be conducted one participant at a time.

After the briefing, you will be given time to study the appropriate area aeronautical charts. All charts will be available for reference for the duration of the study and each pair of origin and destination airports used in the set of clearances will be highlighted, as well. A print out with the Jeppesen Sanderson Instrument Rating Manual's recommended shorthand formatting and symbols will be available for reference during both the training and test trials sessions in case you are not very familiar with that particular method, or have been using an alternative shorthand.

After you have studied the charts, and reviewed the shorthand print out, you will be instructed on how to use the equipment, and will receive approximately 30 minutes of training including eight trial runs, representative of all experimental conditions, in a random order. You will then be asked to begin the test trials.

During the test trials, you will be wearing a headset and will be asked to listen and read back 20 IFR departure and arrival clearances from the Dallas Fort Worth area. A desktop view of the Boeing 777 flight deck from the X-Plane ® Home Edition software will be displayed on a LCD monitor along with a voice-to-text transcription (V-T-T) text strip and a green annunciator light as shown on Figure 1 below.

Each IFR clearance will either have or not have a V-T-T displayed in the text strip (i.e., the V-T-T will be displayed in a random manner). Therefore, you are encouraged to shorthand each clearance on the note pad provided the same way you would during any real flight as the pilot monitoring.

Each clearance will be played back through your headsets and the time intervals between clearances will be between 45 and 120 sec. At the end of each play back, you will be asked to count backwards by 3 from a 3 digit number for about 20 seconds. Then you will be prompted to read back the clearance in its entirety, and it will up to you whether to use (a) your own shorthand, (b) the V-T-T (if available), or (c) your own memory, to do so. However, for this study you will not be able to request the clearance be played back a second time. Do the best you can reading back each clearance in its entirety.

After you complete each read-back, you will be asked to fill out the NASA TLX workload ratings.

Through out the test trials session, you will also be asked to manually record the time in minutes and seconds (from a digital clock which will be provided) when the green annunciator light is ON. The score card with these times will be collected at the end of the test trails session.

The NASA TLX Sources-of-Workload evaluation will be conducted after the test trials are complete. This evaluation is a technique that has been developed by NASA to assess the relative importance of six factors in determining how much workload you experienced while listening and reading back the clearances.

At the end of the test trials a short survey on the Voice-To-Text (V-T-T) transcription usability will be conducted.

Location: To participate in the study, you will come to the Department of Human Factors and Systems on the third floor of Lehman Building located on the campus of Embry-Riddle Aeronautical University.

Time required: We expect that you will be in this research study for only one approximately 2 hour long session.

Audio or video taping: You will be audio taped during this study. If you do not want to be audio taped, you will not be able to participate in the study. Discuss this with the researcher or a research team member. However, to protect your confidentiality, no personally identifiable information will be used during the test trials (e.g., your name) to communicate with you. Upon completion of the research, all audio files will be transcribed and verified to remove any additional personally identifiable information and the audio files will be destroyed within one year after completing the research. Until then, the tape will be kept in a locked, safe place.

Risks: Participation in the study does not involve any risks other than those commonly associated with the use of computer display terminals. Participants will not be exposed to any safety hazards other than those normally encountered in their day-to-day activities. No other physical, psychological, or economic harm is anticipated.

Benefits: We cannot promise any benefits to you or others from your taking part in theis research. However, possible benefits from participation in the study may include improved pilot-controller voice radio communications skills necessary for flying in controlled airspace and under Instrument Flight Rules (IFR).

Compensation or payment: There is no direct compensation for taking part in this study. It is possible, however, that extra credit may be offered for your participation, but this benefit is at the discretion of your instructor.

Study contact for questions about the study or to report a problem: If you have questions, concerns, or complaints, contact Daniela Kratchounova, Principal Investigator, at (386) 763-1969 or by email at kratchounova@knights.ucf.edu, or Dr. Peter Hancock, Faculty Supervisor, Department of Psychology, at 407-823-2310 or by email at Peter.Hancock@ucf.edu, or Dr. Florian Jentsch, Faculty Supervisor, Department of Psychology at (407) 882- 0304 or by email at Florian.Jentsch @ucf.edu.

IRB contact about your rights in the study or to report a complaint: Research at the University of Central Florida involving human participants is carried out under the oversight of the Institutional Review Board (UCF IRB). This research has been reviewed and approved by the IRB. For information about the rights of people who take part in research, please contact: Institutional Review Board, University of Central Florida, Office of Research & Commercialization, 12201 Research

Parkway, Suite 501, Orlando, FL 32826-3246 or by telephone at (407) 823-2901. You may also talk to them for any of the following:

- Your questions, concerns, or complaints are not being answered by the research team.
- You cannot reach the research team.
- You want to talk to someone besides the research team.
- You want to get information or provide input about this research.

Withdrawing from the study: If you decide to leave the study, contact the investigator so that the investigator can terminate the experiment. Credit (if applicable) will still be awarded. The person in charge of the research study can remove you from the research study without your approval. Possible reasons for removal include not meeting the flight hour's requirement. The principal investigator will tell you about any new information that may affect your health, welfare or choice to stay in the research.

Your signature below indicates your permission to take part in this research.

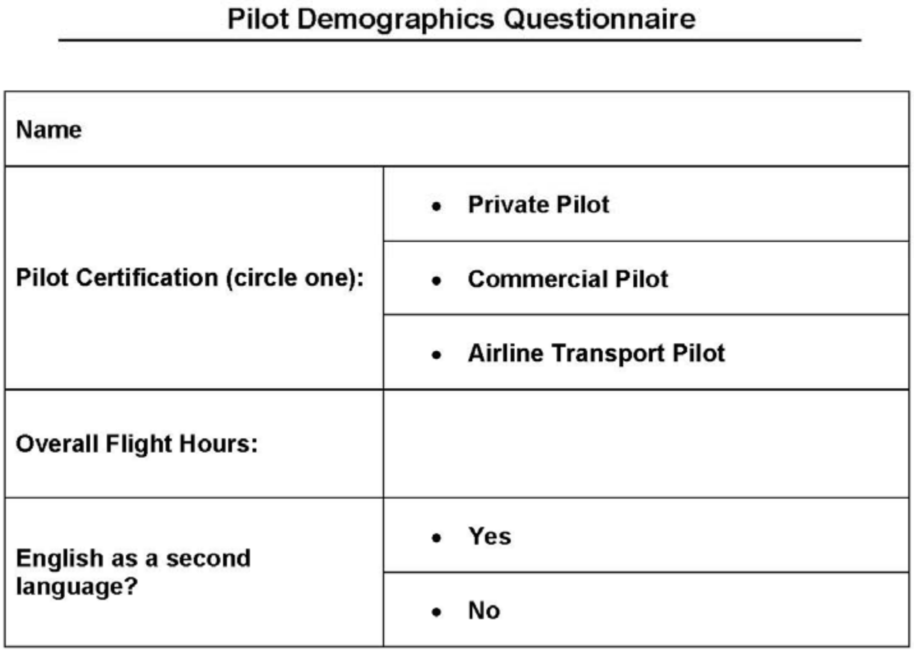
DO NOT SIGN THIS FORM AFTER THE IRB EXPIRATION DATE BELOW

Name of participant

Signature of participant — Date

University of Central Florida IRB
IRB NUMBER: SBE-11-07846
IRB APPROVAL DATE: 10/4/2011
IRB EXPIRATION DATE: 10/3/2012

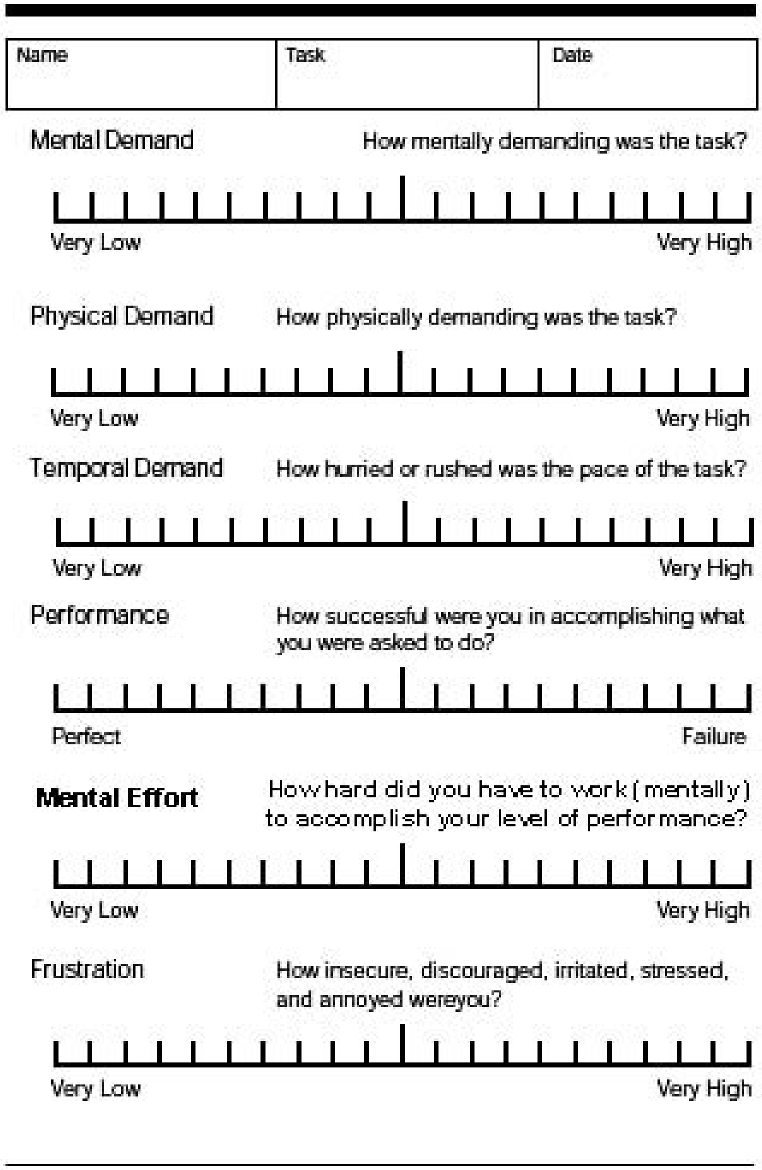
APPENDIX C: DEMOGRAPHICS QUESTIONNAIRE



Pilot Demographics Questionnaire

Name	
Pilot Certification (circle one):	- **Private Pilot**
	- **Commercial Pilot**
	- **Airline Transport Pilot**
Overall Flight Hours:	
English as a second language?	- **Yes**
	- **No**

APPENDIX D: NASA TLX QUESTIONNAIRE



Name	Task	Date

Mental Demand How mentally demanding was the task?

Very Low Very High

Physical Demand How physically demanding was the task?

Very Low Very High

Temporal Demand How hurried or rushed was the pace of the task?

Very Low Very High

Performance How successful were you in accomplishing what you were asked to do?

Perfect Failure

Mental Effort How hard did you have to work (mentally) to accomplish your level of performance?

Very Low Very High

Frustration How insecure, discouraged, irritated, stressed, and annoyed wereyou?

Very Low Very High

APPENDIX E: USABILITY QUESTIONNAIRE

Name: ……………………………………………   **Date:** ………..

"Voice-to-Text" (V-T-T) Transcription Display Usability Survey

Instructions

The purpose of this evaluation is to provide user feedback regarding the usability of V-T-T display utilized in the study:

- Function
- Format (shorthand)
- Ease of Use
- Ease of Interpretation

Function

On a scale of 1 to 5, please give a rating that best represents the quality of V-T-T transcription display (e.g., visibility, legibility, presentation).

1	**2**	**3**	**4**	**5**
Very Poor	*Poor*	*Acceptable*	*Good*	*Very Good*

Format

On a scale of 1 to 5, please give a rating that best represents the quality of V-T-T transcription format (e.g., shorthand).

1	**2**	**3**	**4**	**5**
Very Poor	*Poor*	*Acceptable*	*Good*	*Very Good*

Ease of Use

On a scale of 1 to 5, please give a rating that best represents how easy was to use the V-T-T transcription display.

1	**2**	**3**	**4**	**5**
Very Difficult	*Difficult*	*Neutral*	*Easy*	*Very Easy*

Ease of Interpretation

On a scale of 1 to 5, please give a rating that best represents how easy was to interpret the V-T-T transcription display.

1	**2**	**3**	**4**	**5**
Very Difficult	*Difficult*	*Neutral*	*Easy*	*Very Easy*

Additional Comments:

……………………………………………………………………………………………………

……………………………………………………………………………………………………

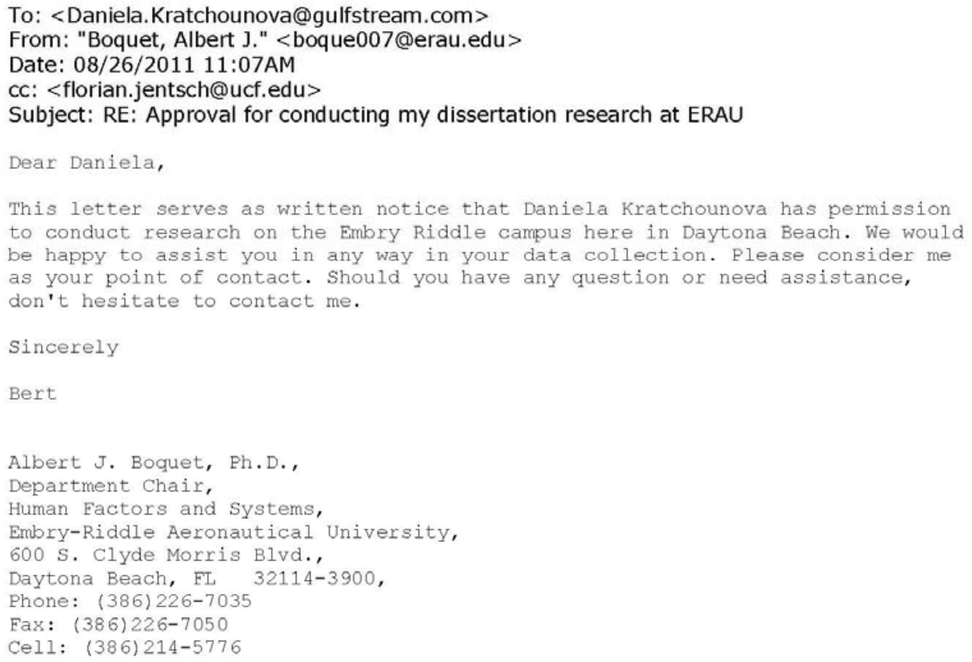
……………………………………………………………………………………………………

……………………………………………………………………………………………………

……………………………………………………………………………………………………

……………………………………………………………………………………………………

APPENDIX F: APPROVAL LETTER TO CONDUCT REASERCH AT EMBRY-RIDDLE AERONAUTICAL UNIVERSITY



To: <Daniela.Kratchounova@gulfstream.com>
From: "Boquet, Albert J." <boque007@erau.edu>
Date: 08/26/2011 11:07AM
cc: <florian.jentsch@ucf.edu>
Subject: RE: Approval for conducting my dissertation research at ERAU

Dear Daniela,

This letter serves as written notice that Daniela Kratchounova has permission to conduct research on the Embry Riddle campus here in Daytona Beach. We would be happy to assist you in any way in your data collection. Please consider me as your point of contact. Should you have any question or need assistance, don't hesitate to contact me.

Sincerely

Bert

Albert J. Boquet, Ph.D.,
Department Chair,
Human Factors and Systems,
Embry-Riddle Aeronautical University,
600 S. Clyde Morris Blvd.,
Daytona Beach, FL   32114-3900,
Phone: (386)226-7035
Fax: (386)226-7050
Cell: (386)214-5776

REFERENCES

Ackerman, P., & Beier, M. (2005, June 1). Age, Ability, and the Role of Prior Knowledge on the Acquisition of New Domain Knowledge: Promising Results in a Real-World Learning Environment. *Psychology and Aging, 20*(2), 341-355.

Ackerman, P., & Beier, M. (2007, December 1). Further explorations of perceptual speed abilities in the context of assessment methods, cognitive abilities, and individual differences during skill acquisition. *Journal of Experimental Psychology: Applied, 13*(4), 249-272.

Ayres, P., & Sweller, J. (2005). The Split-Attention Principle in Multimedia Learning. In R. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 135-146). New York, NY, US: Cambridge University Press.

Baddeley, A. (1981). The concept of working memory: A view of its current state and probable future development. *Cognition, 10*(1-3), 17-23.

Baddeley, A. (1986). *Working memory.* New York, NY, US: Clarendon Press/Oxford University Press.

Baddeley, A. (1992, January 1). Working memory. *Science, 255*(5044), 556-559.

Baddeley, A. (1996). Exploring the central executive. [Special issue]. *He Quarterly Journal of Experimental Psychology A: Human Experimental Psychology, 49A*(10), 5-28.

Baddeley, A. (1998). Recent developments in working memory. *Current Opinion in Neurobiology, 8*(2), 234-238.

Baddeley, A. (2000). Working memory. In A. E. Kazdin (Ed.), *Encyclopedia of psychology* (pp. 276-279). Washington, DC, US: American Psychological Association.

Baddeley, A. (2002). Is working memory still working? *European Psychologist, 7*(2), 85-97.

Baddeley, A. (2003). Working Memory: Looking Back and Looking Forward. *Nature Reviews Neuroscience, 4*(10), 829-839.

Baddeley, A., & Hitch, G. (1994, October 1). Developments in the concept of working memory. *Neuropsychology, 8*(4), 485-493.

Baddeley, A., & Larsen, J. (1994). *Your memory: A user's guide.* New York, NY, US: Penguin Press.

Baddeley, A., & Larsen, J. (2007, April 1). The phonological loop: Some answers and some questions. *The Quarterly Journal of Experimental Psychology, 60*(4), 512-518.

Baron, R. M., & Kenny. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic and statistical considerations. *Journal of Personality and Social Psychology, 51*(6), 1173-1182.

Brown, J. (1958). Some Tests of the Decay Theory of Immediate Memory. *Quarterly Journal of Experimental Psychology, 100,* 12-21.

Brunken, R., Plass, J., & Leutner, D. (2003). Direct measurement of cognitive load in multimedia learning. *Educational Psychologist, 38*(1), 53-61.

Chandler, P., & Sweller, J. (1991). Cognitive load theory and the format of instruction. *Cognition and Instruction, 8*(4), 293-332.

Chandler, P., & Sweller, J. (1996, April 1). Cognitive load while learning to use a computer program. *Applied Cognitive Psychology, 10*(2), 151-170.

Chase, W. G., & Simon, H. A. (1973, January 1). Perception in chess. *Cognitive Psychology, 4*(1), 55-81.

Clark, J. M., & Paivio, A. (1991, September 1). Dual coding theory and education. *Educational Psychology Review, 3*(3), 149-210.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences.* Hillsdale, NJ: Erlbaum.

Collette, F., Salmon, E., Van, D. L., M, Chicherio, C., Belleville, S. et al. (1999, January 1). Regional brain activity during tasks devoted to the central executive of working memory. *Cognitive Brain Research, 7*(3), 411-417.

Farley, T., Hansman, R., Amonlirdviman, K., & Endsley, M. (2000). Shared information between pilots and controllers in tactical air traffic control. *Journal of Guidance, Control and Dynamics, 23*(5), 826-836.

Faul F., E. E., Lamg A-G. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*(2), 175-191.

Flathers II, G. (NASA Contractor). (1987). *Development of an Air Ground Data Exchange Concept: Flight Deck Perspective* (NASA No. NAS1-17974). Langley Research Center.

Gopher, D., & Braune, R. (1984). On the psychophysics of workload: Why bother with subjective measures? [Special issue]. *Human Factors, 26*(5), 519-532.

Graff, M. (2005, April 1). Individual differences in hypertext browsing strategies. *Behaviour & Information Technology, 24*(2), 93-99.

Guilford, J. P., & Zimmerman, W. S. (1981). *The Guilford-Zimmerman aptitude survey: Manual of instructions and interpretations.* Palo Alto: Consulting Psychologists Press.

Hancock, P. A., Williams, G., Manning, C. M., & Miyake, S. (1995). Influence of task demand characteristics on workload and performance. *International Journal of Aviation*

Psychology, 5(1), 63-86.

Harp, S., & Mayer, R. (1998, September 1). How seductive details do their damage: A theory of cognitive interest in science learning. *Journal of Educational Psychology, 90*(3), 414-434.

Hart, S. G. (2006). NASA-Task Load Index (NASA-TLX); 20 Years Later*Proceedings of the Human Factors and Ergonomics Society 50th Annual Meeting* (pp. 904-908). Santa Monica.

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

Helleberg, J., & Wickens, C. (2003, July 1). Effects of Data-Link Modality and Display Redundancy on Pilot Performance: An Attentional Perspective. *International Journal of Aviation Psychology, 13*(3), 189-210.

Jeung, H., Chandler, P., & Sweller, J. (1997, September 1). The role of visual indicators in dual sensory mode instruction. *Educational Psychology, 17*(3), 329-343.

Judd, C., Kenny, D., McClelland, G., &. (2001, June). Estimating and testing mediation and moderation in within-subject designs. *Psychological Methods, 6*(2), 115-134.

Kalyuga, S. (2005). Prior Knowledge Principle in Multimedia Learning. In R. Mayer (Ed.), *The Cambridge handbook of multimedia learning.* (pp. 325-337). New York, NY, US: Cambridge University Press.

Kalyuga, S., Ayres, P., Chandler, P., & Sweller, J. (2003). The expertise reversal effect. *Educational Psychologist, 38*(1), 23-31.

Kalyuga, S., Chandler, P., & Sweller, J. (1998). Levels of expertise and instructional design. *Human Factors, 40*(1), 1-17.

Kalyuga, S., Chandler, P., & Sweller, J. (1999, August 1). Managing split-attention and redundancy in multimedia instruction. *Applied Cognitive Psychology, 13*(4), 351-371.

Kalyuga, S., Chandler, P., & Sweller, J. (2000). Incorporating learner experience into the design of multimedia. *Journal of Educational Psychology, 92,* 126-136.

Kalyuga, S., Chandler, P., & Sweller, J. (2004). When redundant on-Screen text in multimedia technical instruction can interfere with learning. *Human Factors, 46*(3), 567-581.

Kerns, K. (1991). Data-link communication between controllers and pilots: A review and synthesis of the simulation literature. *International Journal of Aviation Psychology, 1*(3), 181-204.

Lancaster, J., & Casali, J. (2005, September 1). Investigating pilot performance using mixed-modality simulated data link. *In Proceedings of the Human Factors and Ergonomics Society 49th Annual Meeting, 1,* 188-122.

Larsen, J. D., Baddeley, A., & Andrade, J. (2000, May). Phonological similarity and the irrelevant speech effect: Implications for models of short-term verbal memory. *Memory, 8*(3), 145-158.

Leahy, W., Chandler, P., & Sweller, J. (2003, May). When auditory presentations should and should not be a component of multimedia instruction. *Applied Cognitive Psychology, 17*(4), 401-418.

Leutner, D., & Plass, J. (1998, December 1). Measuring learning styles with questionnaires versus direct observation of preferential choice behavior in authentic learning situations:

The Visualizer/Verbalizer Behavior Observation Scale (VV-BOS). *Computers in Human Behavior, 14*(4), 543-557.

Levie, W., & Lentz, R. (1982, Winter). Effects of text illustrations: A review of research. *Educational Communication & Technology Journal, 30*(4), 195-232.

Loftus, G., Dark, V., & Williams, D. (1979, April 1). Short-term memory factors in ground controller/pilot communication. *Human Factors, 21*(2), 169-181.

Low, R., & Sweller, J. (2005). The Modality Principle in Multimedia Learning. In R. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 147-158). New York, NY, US: Cambridge University Press.

Marcus, N., Cooper, M., & Sweller, J. (1996, March 1). Understanding instructions. *Journal of Educational Psychology, 88*(1), 49-63.

Mayer, R. (1997, Winter). Multimedia learning: Are we asking the right questions? *Educational Psychologist, 32*(1), 1-19.

Mayer, R. (2001, June 1). What good is educational psychology? The case of cognition and instruction. *Educational Psychologist, 36*(2), 83-88.

Mayer, R. (2005). Cognitive Theory of Multimedia Learning. In R. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 31-48). New York, NY, US: Cambridge University Press.

Mayer, R. (2005). Introduction to multimedia learning. In R. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 1-16). New York, NY, US: Cambridge University Press.

Mayer, R. (2005). Principles for Reducing Extraneous Processing in Multimedia Learning:

Coherence, Signaling, Redundancy, Spatial Contiguity and Temporal Contiguity Principles. In R. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 183-200). New York, NY, US: Cambridge University Press.

Mayer, R., & Anderson, R. (1991, December 1). Animations need narrations: An experimental test of a dual-coding hypothesis. *Journal of Educational Psychology, 83*(4), 484-490.

Mayer, R., & Anderson, R. (1992, December 1). The instructive animation: Helping students build connections between words and pictures in multimedia learning. *Journal of Educational Psychology, 84*(4), 444-452.

Mayer, R., Bove, W., Bryman, A., Mars, R., & Tapangco, L. (1996, March 1). When less is more: Meaningful learning from visual and verbal summaries of science textbook lessons. *Journal of Educational Psychology, 88*(1), 64-73.

Mayer, R., & Gallini, J. (1990, December 1). When is an illustration worth ten thousand words? *Journal of Educational Psychology, 82*(4), 715-726.

Mayer, R., Heiser, J., & Lonn, S. (2001). Cognitive constraints on multimedia learning: When presenting more material results in less understanding. *Journal of Educational Psychology, 93*(1), 187-198.

Mayer, R., & Johnson, C. (2008, May). Revising the redundancy principle in multimedia learning. *Journal of Educational Psychology, 100*(2), 380-386.

Mayer, R., & Moreno, R. (1998, June 1). A split-attention effect in multimedia learning: Evidence for dual processing systems in working memory. *Journal of Educational Psychology, 90*(2), 312-320.

Mayer, R., & Moreno, R. (2003). Nine ways to reduce cognitive load in multimedia learning.

Educational Psychologist, 38(1), 43-52.

Mayer, R., & Sims, V. (1994, September 1). For whom is a picture worth a thousand words? Extensions of a dual-coding theory of multimedia learning. *Journal of Educational Psychology, 86*(3), 389-401.

McGann, A., Morrow, D., Rodvold, M., & Mackintosh, M. (1998). Mixed-media communication on the flight deck: A comparison of voice, data link, and mixed ATC environments. *International Journal of Aviation Psychology, 8*(2), 137-156.

Merrill, M. D. (2002). First principles of instruction. *Educational Technology Research and Development, 50*(3), 43-59.

Moreno, R., & Mayer, R. (1999, June 1). Cognitive principles of multimedia learning: The role of modality and contiguity. *Journal of Educational Psychology, 91*(2), 358-368.

Moreno, R., & Mayer, R. (2002, March 1). Verbal redundancy in multimedia learning: When reading helps listening. *Journal of Educational Psychology, 94*(1), 156-163.

Moreno, R., & Plass, J. (2006). Individual Differences in Learning with Verbal and Visual Representations. In *NYU Symposium on Technology and Learning.*

Mousavi, S. Y., Low, R., & Sweller, J. (1995). Reducing cognitive load by mixing auditory and visual presentation modes. *Journal of Educational Psychology, 87*(2), 319-334.

Muller, D., Judd, C. M., & Yzerbyt, V. Y. (2005, December 1). When moderation is mediated and mediation is moderated. *Journal of Personality and Social Psychology, 89*(6), 852-863.

Neath, I. (2000, September 1). Modeling the effects of irrelevant speech on memory. *Psychonomic Bulletin & Review, 7*(3), 403-423.

Paas, F. (1992, December 1). Training strategies for attaining transfer of problem-solving skill in statistics: A cognitive-load approach. *Journal of Educational Psychology, 84*(4), 429-434.

Paas, F., Renkl, A., & Sweller, J. (2003). Cognitive load theory and instructional design: Recent developments. *Educational Psychologist, 38*(1), 1-4.

Paas, F., Tuovinen, J., Tabbers, H., & van Gerven, P. (2003). Cognitive load measurement as a means to advance cognitive load theory. *Educational Psychologist, 38*(1), 63-71.

Paas, F., & van Merrienboer, J. (1994, December 1). Instructional control of cognitive load in the training of complex cognitive tasks. *Educational Psychology Review, 6*(4), 351-371.

Paas, F., & van Merrienboer, J. (1994). Measurement of cognitive load in instructional research [Special issue]. *Perceptual and Motor Skills, 79,* 419-430.

Paivio, A. (1986). *Mental representations.* New York, NY, US: Oxford University Press.

Paivio, A. (1991, September 1). Dual coding theory: Retrospect and current status. *Canadian Journal of Psychology, 45*(3), 255-287.

Penney, C. G. (1989). Modality effects and the structure of short-term verbal memory. *Memory and Cognition, 17,* 398-422.

Peterson, L. R., & Peterson, M. J. (1959). Short-term retention of individual verbal items. *Journal of Experimental Psychology, 59,* 193-198.

Plass, J., Chun, D., Mayer, R., & Leutner, D. (2003, March 1). Cognitive load in reading a foreign language text with multimedia aids and the influence of verbal and spatial abilities. *Computers in Human Behavior, 19*(2), 221-243.

Pollock, E., Chandler, P., & Sweller, J. (2002). Assimilating complex information [Special issue]. *Learning and Instruction, 12*(1), 61-86.

Rehmann, A. (1995). *Flight Simulator Evaluation Of Baseline Crew Performance With Three Data Link Interfaces* (Tech. Rep. No. DOT/FAA/CT-TN95/19).

Rehmann, A. (1995, December 1). *Handbook of human performance measures and crew requirements for flight deck research* (Crew System Ergonomics Information Analysis Center (CSERIAC) No. DOT/FAA/CT-TN95/49). US DOT FAA Technical Center Atlantic City International Airport, NJ 08405.

Rehmann, A. (1997). *Human Factors Recommendations for Airborne Controller-Pilot Data Link Communications (CPDLC) Systems: A Synthesis of Research Results and Literature* (Tech. Rep. No. DOR/FAA/CT-TN97/6).

Sadler-Smith, E., & Smith, P. (2004, July 1). Strategies for accommodating individuals' styles and preferences inflexible learning programmes. *British Journal of Educational Technology, 35*(4), 395-412.

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science, 12,* 257-285.

Sweller, J. (1994). Cognitive load theory, learning difficulty, and instructional design. *Learning and Instruction, 4,* 295-312.

Sweller, J. (2005). Implications of Cognitive Load Theory for Multimedia Learning. In R. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 19-30). New York, NY, US: Cambridge University Press.

Sweller, J. (2005). The Redundancy Principle in Multimedia Learning. In R. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 159-167). New York, NY, US:

Cambridge University Press.

Sweller, J. (2010). Element interactivity and intrinsic, extraneous, and germane cognitive load. *Educational Psychology Review, 22*(2), 123-138.

Sweller, J., & Chandler, P. (1994). Why some material is difficult to learn. *Cognition and Instruction, 12*(4), 185-233.

Sweller, J., van Merriлnboer J., & Paas, F. (1998, September 1). Cognitive architecture and instructional design. *Educational Psychology Review, 10*(3), 251-296.

Tabbers, H. K. (2002). *The modality of text in multimedia instructions: Refining the design guidelines*, Open University of the Netherlands.

Tabbers, H., Martens, R., & van Merriлnboer, J. (2004, March 1). Multimedia instructions and cognitive load theory: Effects of modality and cueing. *British Journal of Educational Psychology, 74*(1), 71-81.

Tindall-Ford, S., Chandler, P., & Sweller, J. (1997). When two sensory modes are better than one. *Journal of Experimental Psychology: Applied, 3*(4), 257-287.

Tulving, E. (1986, September 1). Episodic and semantic memory: Where should we go from here? *Behavioral and Brain Sciences, 9*(3), 573-577.

Tulving, E. (1993, June 1). What is episodic memory? *Current Directions in Psychological Science, 2*(3), 67-70.

Van Merriлnboer, J., & Sweller, J. (2005, June 1). Cognitive Load Theory and Complex Learning: Recent Developments and Future Directions. *Educational Psychology Review, 17*(2), 147-177.

Wickens, C., Goh, J., Helleberg, J., & Talleur, D. (2002). *Modality differences in advanced* *cockpit displays: Comparing auditory vision and redundancy for navigational communications and traffic awareness* (NASA No. ARL-02-8/NASA-02-6). Ames Research Center Moffett Field, CA.

Wiegmann, D., & Shappell, S. (2001). Human error perspectives in aviation. *International Journal of Aviation Psychology, 11*(4), 341-357.

Wittrock, M. C. (1989, Fall). Generative processes of comprehension. *Educational Psychologist, 24*(4), 345-376.

Xie, B., & Salvendy, G. (2000). Prediction of mental workload in single and multiple tasks environments. *International Journal of Cognitive Ergonomics, 4*(3), 213-242.