The Pennsylvania State University

The Graduate School

College of the Liberal Arts

TASK AND INDIVIDUAL CHARACTERISTICS AS PREDICTORS OF PERFORMANCE IN A JOB-RELEVANT MULTI-TASKING ENVIRONMENT

A Thesis in

Psychology

by

Theodore B. Kinney

Submitted in Partial Fulfillment
of the Requirements
for the Degree of

Doctor of Philosophy

August 2007

The thesis of Theodore B. Kinney was reviewed and approved* by the following:

Kevin R. Murphy
Professor of Psychology
Thesis Adviser
Chair of Committee

Jeanette Cleveland
Professor of Psychology

James L. Farr
Professor of Psychology

Michael J. Wenger
Associate Professor of Psychology

Frank E. Ritter
Associate Professor of Information Sciences Technology

Melvin M. Mark
Professor of Psychology
Head of the Department of Psychology

*Signatures are on file in the Graduate School

Abstract

Understanding performance in complex multitasking environments is of paramount importance to selection practitioners. Organizations are asking employees to do more work in less time than ever before. With this increased task demand, understanding who can and cannot survive in this type of situation is critical. To date, little research has investigated how to predict who will succeed and who will fail in these demanding situations. Cognitive psychological research has focused on task characteristics that impact human capacity to multitask. This stream of research has not had much impact on the selection practitioner. Research on individual difference predictors of job performance has ignored criteria specific to performance in multitasking job environments. This dissertation blends what cognitive psychologists have learned about human capacity to multitask and what industrial/organizational psychologists have learned about predicting job performance to investigate individual difference predictors in a job-relevant multitasking simulation. The results from this study suggest that several of the task characteristic conditions researched in cognitive multitasking research do generalize to applied multitasking environments. Also, several individual difference variables (cognitive ability, intellectance, extroversion, and polychronicity) emerged as predictors of performance in an applied multitasking environment, with cognitive ability being the best predictor. Implications of these results on employee selection, job design, and future research are discussed.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

Jobs in modern organizations are more complex than ever. In these fast-paced, demanding, multiple-task-goal environments, interest in multitasking is on the upswing. Managers in organizations and the popular press alike discuss "multitasking ability" as an important characteristic of successful job performers (e.g., Santosus, 2003; Bremner, 2005). Santosus (2003) discussed how "multitasking – both in the sense of doing more than one task at a time as well as switching among tasks – has taken on added importance to companies" (pg. 1). This author also stated that organizations have become chronically understaffed creating the need for increased focus on efficiency and the ability to "juggle as many jobs as one can." In a recent Wall Street Journal article Sandberg (2006) wrote that "Today's workers typically have to do three things while simultaneously arguing on the phone with a spouse."

The scientific community is also recognizing the importance of multitasking in today's jobs (e.g., Joslyn & Hunt, 1998; Ishizaka, Marshall, & Conte, 2001; Konig, Buhner, & Murling, 2005). Joslyn and Hunt (1998) pointed out that the ability to make decisions within a few seconds to a couple of minutes is the hallmark of many applied situations. The need to predict who will be able to survive in a multitasking-rich environment has become of critical importance to hiring managers tasked with the need to select people into these high-multitasking demand positions.

The need for multitasking in the workforce is not a new idea. Mintzberg (1973) identified several aspects of managerial performance that suggest a need to multitask. Mintzberg explained that performance in managerial positions is characterized by"brevity, variety, and fragmentation" (pg. 31). As will be discussed later, these descriptors are critical components of a multitasking environment.

Since Mintzberg's work in the Seventies, jobs have become increasingly complex and the need to multitask has grown beyond managerial jobs. Many entry-level jobs of today have an amazing degree of multitasking required to be successful. Modern call center employees are a prime example (Kinney, Phillips, Michel, and Higgins, 2005). As observed in a series of recent job analyses conducted at 10 large call centers across four different job functions, entry-level employees in call centers were often required to manage many tasks in the same general time period and frequently switch focus and reprioritize tasks (Kinney, et al 2005). In order to process a single customer inquiry, these call center associates often have to talk on the phone while manipulating complex computer systems, writing and reading emails, and documenting the process as they go. In these jobs "efficiency" or "handling time" are key metrics that define performance, raises, promotions, etc. Simply put, if a call center employee cannot multitask, a call center employee cannot survive in that environment (Kinney et al, 2005).

While the need to understand performance in applied multitasking environments is clear, research on the topic of multitasking has traditionally emerged out of the highly controlled laboratories of cognitive and experimental psychologists. The goal of these research streams has been to identify similarities across people, for example, in human capacity to process multiple pieces of information at a single point in time (Delbridge, 2000). This research has been insightful in understanding human information processing mechanisms; however, for the selection practitioner with the goal of identifying meaningful individual differences to predict performance in environments with a high need for multitasking, this line of research has limited usefulness. Thus, selection practitioners, who need to discriminate between those who have the potential to succeed in a multitasking environment and those who cannot, can gain limited insight from existing research on the topic. With relatively few exceptions (e.g., Delbridge, 2000; Ishizaka et al, 2001; Konig et al, 2006; Zaldivar, 2006; Russell & Sacco, 2006), the selection practitioner has little guidance available from the research community.

In light of this disconnect between science and practice on the topic of selecting employees into a high multitasking environment, this thesis takes steps toward bridging this growing and critical gap. Specifically, the purpose of this dissertation is to explore individual differences antecedent to performance in a job relevant multitasking environment. If meaningful individual differences are discovered, the study can provide guidance on how selection practitioners should go about understanding predictors used to select employees into multitasking environments. Further, this study will extend the findings of multitasking lab research into an applied job-relevant environment. A model of the two general purposes of this study is presented in Figure 1.

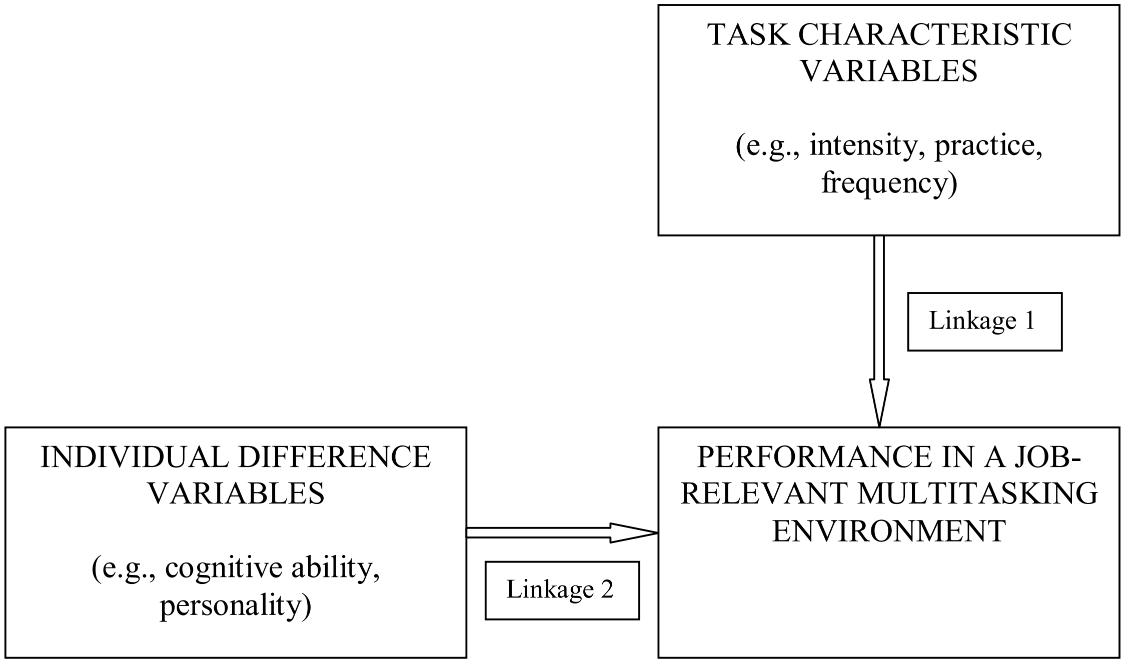

Figure 1. General Conceptual Model under Study in the Thesis.

Cognitive and Experimental Research on Multitasking

As mentioned earlier, the bulk of the scientific multitasking research has been aimed at understanding trends in the human capacity to multitask under different task conditions rather than to understand antecedents of performance differences in multitasking environments. These research streams have traditionally focused on basic tasks, such as discriminating between even and odd numbers, which are not relevant to most work situations (e.g., Guzy & Axelrod, 1972; Laabs & Stanger, 1976; Weber, Blagowski, & Mankin, 1982; Weber, Burt, & Noll, 1986; Rogers & Monsell, 1995; Gopher, Armony, & Greenspan, 2000). Gopher and Kahneman (1971) stated that research has focused on performance in multitasking environments, but has not focused on individual differences that impact performance. This statement still applies 35 years later. Further, much of the multitasking research has taken a within subjects approach to identify differences across differing task characteristics; whereas, what is needed for selection practitioners is an approach that looks between subjects and at how individuals differ in their capacity to function in multitasking environments. This cognitive and experimental research has primarily focused on research questions relevant to Linkage 1 in Figure 1.

Because of these challenges to the external validity of most of the basic lab research, these studies cannot serve as "manuals" for a hiring manager to follow to select associates, but many relevant lessons for understanding how multitasking environments impact performance can be learned from this stream of research. As Weber, Burt, and Noll (1986) stated, many applied tasks involve similar processes to those observed in the lab setting. Specifically, these authors stated that common tasks involve a rapid cycling of attention or switching between information available in the environment and information stored in memory. To the extent that these common tasks involve the same types of cognitive processing studied in the lab, this lab research can provide a framework for understanding the key features of a multitasking environment that can impact performance (e.g., Rogers & Monsell, 1995). This lab research has typically focused on two broad topics when investigating human capacity to multitask: Dual task performance and task switching.

Dual Task Research

The nature of dual-task performance research has been to investigate the most basic definition of multitasking: the capacity for people to do multiple tasks at the same time (Delbridge, 2000). The typical procedure followed in these studies is to have participants focus on the performance of one task while maintaining performance on a secondary task. Delbridge (2000) explains that this methodology is employed to evaluate the use and distribution of attentional resources. If the participant does not have enough attentional resources available to maintain the secondary task, performance deficiencies appear (Meyer & Kieras, 1997). The approach to understanding dual-task performance is based on the belief that humans have a finite capacity for attention. The primary finding to come out of this line of investigation is that task performance is generally negatively impacted if attentional resources must be shared with a second simultaneous task (e.g., Damos & Smist, 1982). Examples of these studies are reviewed below.

Lindenberger, Marsiske, and Baltes (2000) showed that performing a challenging walking task while simultaneously performing a memory task becomes increasingly difficult across age groups. The theory behind this study was that as adults advance in age, more attentional resource must be allocated to locomotion. Thus, older adults completing a secondary memory task have more demands competing for finite attentional resource than do younger adults.

The memory task involved listening to sets of words delivered ten seconds apart and then typing the words into a computer at the end of the encoding stage. Performance was operationalized as the number of words correct. The walking task had two conditions. One condition involved walking a relatively simple oval track. The other condition involved walking a complex track with 22 turning points at various angles. In the dual task trials, the participants walked a track during the encoding stage of the memory task, and then were seated at a computer for the recall portion of the memory task.

Lindenberger et al. (2000) found that, regardless of the characteristics of the task, younger participants out-performed older participants. Further, the mean performance difference in recall with seated encoding and recall with walking encoding was greater for the older group than the younger group. This pattern of results suggested that individual differences in age predicted dual task costs.

The work of Lindenberger et al. (2000) is notable on a couple levels. First, the tasks involved in the study were intentionally more complex than most multitasking lab studies. While encoding while walking a track is not immediately generalizable to most jobs, this slightly more complex task relative to traditional dichotic listening tasks shows that multitasking costs are observed at varying levels of task complexity. Second, this study was focused on the impact of various task characteristics on performance, but also the researchers investigated the impact of an individual difference variable on performance. This study illustrates the potential for individual differences to lead to performance differences in multitasking environments.

Hein and Schubert (2004) extended the findings of Lindenberger et al. (2000) by investigating the underlying source of the observed influence of age on dual task performance. Specifically, these authors used a psychological refractory period (PRP) approach to measure dual-task costs. The PRP approach measures reaction times to tasks separated by variable intervals. In this case, the tasks involved discrimination between high and low tones and differentiating between different letters on a computer screen. Hein and Schubert (2004) found that dual task performance differences between old and young participants could be explained by the older participants' increased susceptibility to input interference. In other words, individual differences in ability to process stimulus material led to meaningful differences in dual task costs.

Task-Switching Research

Research on effects on performance when participants have to switch tasks has been another heavily researched area in the cognitive multitasking literature. In these studies, multitasking is typically operationalized as performance on a simple task with periodic switches to another simple task. In 1971, Gopher and Kahneman used what has become a typical task-switching methodology. These researchers presented participants with different auditory information in each ear. The participants were asked to monitor the information in one ear while ignoring the information presented in the other ear. Periodically, the participants were cued to switch the side that they were to attend. In general, these task switching studies find that performance decreases as the demand for task switching increases. This methodology is referred to as a "dichotic listening task" and was first used by Broadbent (1958).

Salthouse, Fristoe, McGuthry, and Hambrick (1998) employed another typical procedure used to measure task-switching performance. These researchers presented series of numbers to participants and asked them to make a simple decision about the stimuli each time it arrived. For example, they asked participants if the number was even or odd. After several stimuli were presented, the researchers presented a cue to switch tasks and continued presenting stimuli. For example, rather than discriminating between even and odd, the participants had to switch to identifying whether or not the number was greater than or less than 5. Salthouse et al. (1998) then measured the impact on the decisions made immediately following the switch. This general type of task switching methodology is referred to as "list processing" and was pioneered by Jerslid (1927). Studies using this methodology typically observe decrements in performance in task switching environments that are associated with a need to cognitively reconfigure in order to make the switch (Rogers & Monsell, 1995).

Yeung and Monsell (2003) investigated the effect of task set priming on task switching performance. In other words, they were seeking to understand whether or not previous exposures to task sets would lead to increased performance in future trials. To isolate the effects of controlled task exposure on task switching costs, Yeung and Monsell (2003) chose an alphabet-numeric task. The participants were presented with a letter and a number. The participants had to add the amount of letters corresponding to the number. For example, if the researchers presented "H3", the correct participant response is "K". The secondary task was to indicate whether or not the number and the letter correspond in terms of curvature; straight characters contain no curved lines and curved characters contain at least one curve. To complete this task, participants had to state "yes" if the two characters were consistent in curvature (either both have curves or both do not) or "no" if the characters were not consistent in curvature.

The primary task by itself had been shown in previous studies to improve significantly with practice (Klapp, Boches, Trabert, & Logan, 1991; Logan & Klapp, 1991; Logan, 1988). Yeung and Monsell (2003) sought to understand if the practice effects would remain consistent with past research if a secondary task were added. These authors reported that despite practice on both primary and secondary tasks, task switching has associated costs on performance. Regardless of how recent the practice is to the task and regardless of whether or not one or both tasks were practiced, the task switch impacted performance.

These results show the serious impact of task switches on performance. No matter how much experience a person has on a task or set of tasks, an environment involving task switching will lead to differences in performance compared to a single-task environment.

This finding may be relevant to organizations in that most selection components look at single characteristics (e.g., selling skill, typing speed, etc.), but do not consider performance in these skill areas when a person must alternate between them. Because the performance decrement due to task switching is present despite experience, organizations may be better served by understanding individual differences that predict reduced performance decrements than they are by understanding singular skills in a "monotask" environment.

Unfortunately, Yeung and Monsell (2003) did not consider the role of individual difference variables in tasks switching costs. Also, the nature of the tasks was quite different from tasks typically completed in most jobs. Thus, future research is needed to investigate whether or not these findings generalize to an applied organization setting.

Rogers and Monsell (1995) used a similar approach to Yeung and Monsell (2003) to identify the costs of switching task sets in predictable settings. In this study, the control group completed the same task set repeatedly. The experimental group was asked to complete alternating task sets in successive trials. The tasks involved classifying letters as consonants or vowels and classifying numbers as even or odd. In each stimulus presentation, the participant was shown two characters: a number and a letter. If the task was to differentiate between vowels and consonants, the participant was to press the appropriate key relative to the letter in the character pair. If the task was to differentiate between evens and odds, the participant was to press the appropriate key relative to the number in the character pair. The hypothesis tested by the authors was that switching task sets required a cognitive reconfiguration, which would result in a performance decrement.

Rogers and Monsell (1995) found that performance deficiencies were observed both in terms of reaction times and accuracy. Further, the authors found that the cost on performance increased as the need to switch tasks became quicker (i.e., response-stimulus interval became shorter). The authors note that the cost remained relatively consistent across trials despite practice.

The experiments that Rogers and Monsell (1995) conducted show that the cost of task switching is robust. Task switching environments impact performance even when the tasks are non-complex, predictable, and over learned. In organizations, many jobs require lots of task switching on complex tasks that are unpredictable and oftentimes novel or unique: Thus, the impact of task switches observed by these authors is likely to be even greater in applied organizational settings.

Mieran (1996) applied a similar approach to studying performance decrements in task switching situations (this approach has come to be called the "mixed block paradigm"). The difference in Mieran's (1996) study was that the author made the task switches unpredictable. The cue to switch tasks was presented in a random order. The author found that the performance decrements due to task switching were present in this condition as well. Further, the author found that manipulating the response-stimulus interval led to greatly impacted performance. This research suggests that organizations with fast-paced and unpredictable environments should expect performance to be effected relative to performance in a predictable and manageable environment.

Gopher, Armony, and Greenspan (2000) acknowledged that performance levels decrease in a task-switching environment and sought to understand more about the nature of the decrease. Specifically, they looked at the effect on switching costs when participants focused on either speed or accuracy when switching. The authors also looked at the impact of practice on the task performance after a switch.

In these experiments (Gopher et al., 2000), participants were presented with a row of numbers that participants either had to count or add together. The participants were instructed to adopt one of three performance strategies: (a) emphasis on response speed, (b) emphasis on response accuracy, or (c) attribute equal importance to speed and accuracy. These researchers found that any switch, whether it is in the characteristics of the task or the strategy employed to complete it, leads to lower task performance. Further, they found that practice did not significantly reduce the extent to which performance was adversely effected.

Weber, Blagowski, and Mankin (1982) conducted an early task switching study. These authors used a simple but effective task switching procedure to investigate whether or not the act of switching tasks in the simplest of situations impacted performance. The manipulation was response intensity in voicing lists of letters and numbers. For example, if a participant mouths the alphabet with no voice as fast as possible and then does it again in a speaking voice, the time it takes to complete these tasks is similar. If the participant is asked to switch response intensity, the response time is impacted greatly (e.g., a B c D e F g . . . . Z; where lower case = no voice and upper case = speaking voice). In a series of experiments these authors showed slight changes to the characteristics of the task create dramatic decreases in performance. The authors also had the participants repeat these trials over a ten day period and the results remained relatively the same. Practice did not make the task switching performance decrement disappear.

A few years later, Weber and a different group of colleagues (Weber, Burt, & Noll, 1986) employed a similar methodology, but this time they staggered whether the response was generated from memory or from perceptual information. This study found that switching between memory and perception involves using attentional resources. Use of these finite resources can lead to a decrease in performance in a heavy task-switching environment.

In another early test of the impact of task switching on performance, Guzy and Axelrod (1972) presented participants with trains of "clicks" in headphones. The researchers found that participants were able to count clicks more accurately if clicks were delivered in one ear. When the clicks alternated between ears, the accuracy of click counting suffered. The authors concluded that counting across ears involves a shift of attention. The task of shifting attention takes time and resource and, thus, performance decrements were observed.

Effects on performance in these studies are typically operationalized in two ways: performance speed and performance accuracy. In other words, these researchers are interested in reaction times of participants in attending to the cue to switch tasks and the error rate on the task after the switch. Often, both criteria are used and they are highly correlated with each other (Hazeltine, Teague, & Ivry, 2002; Sohn & Anderson, 2001).

Research on the effect of practice in both the dual-task and the task switching paradigms has been mixed. The premise behind this research is that the more practice a participant has in multitasking situations, the less their performance will be impacted in future similar multitasking situations. Van Selst et al. (1999) used a dual task procedure involving visual and auditory tasks and looked at performance over time. The participants' ability to perform both tasks simultaneously improved with practice. The findings in this study contradict the findings of research reviewed earlier, which suggested that practice did not have a significant effect on task switching costs (e.g., Gopher et al., 2000). Hazeltine et al. (2002) found similar results to Van Selst (1999) using a similar procedure. In this case, the researchers had the participants practice the dual-task until their performance in the multitasking condition was indistinguishable from their single task performance in a non-multitasking situation. The impact of practice on performance in environments characterized by demand to switch tasks needs further research.

This review of the cognitive multitasking literature leads to two key conclusions and two key gaps in the research. The thorough and rigorous methodology used in these lab studies has shown that (1) Performance is adversely affected in environments with a high demand for multitasking, and (2) these performance decrements are largely robust across task characteristics, such as, intensity, frequency, practice, experience, and complexity. This research underscores the importance for organizations to be concerned about the nature of the multitasking demands on their jobs. However, this research does not offer insight into (a) how these findings generalize to real world work settings (see Linkage 1 in Figure 1), and (b) how individuals differ in their ability to handle demanding multitasking environments (see Linkage 2 in Figure 1). For organizations to manage to the high multitasking demands in their jobs, these research gaps must be considered.

Individual Differences and Applied Cognitive Research on Multitasking

While task-switching research has not commonly been applied to selection settings, it has translated into several practical applications, some of which begins to address the two critical gaps. The relatively small body of research relevant to understanding how these multitasking findings generalize to applied settings and how individuals vary in their ability to perform in multitasking environments are discussed below.

Age-Related Multitasking Individual Differences

"In recent years cognitive aging researchers have become increasingly interested in whether age-related differences in executive functioning may contribute to an understanding in age-related decline in various intellectual abilities" (Kray, Li, & Lindenberger, 2002, pg. 363). This stream of research is relevant to the research question asked here as the studies in this area have focused on identifying individual difference predictors of performance in multitasking environments. This research in concert with research on driving and flying, to be discussed later, goes beyond the typical multitasking research paradigm employed when measuring the impact of different task characteristics on performance in a multitasking environment. These lines of research introduce the possibility that characteristics of the individual, rather than the task, can be used to predict performance in multitasking environments.

Salthouse et al. (1998) investigated whether or not individual differences in age-related cognitive and perceptual abilities are a critical factor in understanding individual differences in performance in multitasking environments. Further, the authors measured the role of intelligence and working memory to further understand individual difference predictors of multitasking performance.

These researchers found strong evidence for individual differences in task switching performance, as operationalized by measures of reaction times in completing the keying tasks. Inductive reasoning was related to the task switching variables with magnitudes ranging from $r = .31$ to $r = .56$. Spatial ability predicted task switching performance with magnitudes ranging from $r = .19$ to $r = .30$. Working memory predicted task switching with magnitudes ranging from $r = .37$ to $r = .59$. Age predicted task switching performance with magnitudes ranging from $r = .36$ to $r = .64$.

Salthouse et al. (1998) concluded that their observed pattern of results suggests that there are reliable predictors of individual differences in task switching performance. The authors note that these differences were manifested both in terms of “ease” and “efficiency”, which correspond to speed and accuracy measures of task switching performance.

This study is one of the most comprehensive investigations into individual differences in multitasking research to date. While the tasks that the participants completed have little ecological validity for organizations, the tasks were complex and not completely dissimilar to the types of routine tasks that are completed in some jobs. It is encouraging that these authors found a relationship between measures common to selection settings and individual differences in tasks switching performance.

While the tasks used by Salthouse et al. (1998) are more complex than the tasks employed by many of the researchers discussed earlier, there are still some serious threats to the external validity of the findings into an organizational context. On such threat involves the nature of the task switch. In the Salthouse et al. (1998) and many other task switching studies, the switch comes at a time defined by and controlled by the researcher. Further, the number of tasks to switch to is typically pre-defined and limited. Kray, Li, and Lindenberger (2002) attempted to study multitasking performance in an environment more similar to applied settings.

Kray et al. (2002) investigated switching costs when an external switching cue is present in the environment rather than mandated by the researcher and when many potential tasks are available to switch to. These researchers were interested in whether or not age predicted between-person differences in task switching performance.

The authors found that the task switching uncertainty created from using an external switch cue procedure led to drastically increased task switching performance decrements relative to single task performance. These findings suggest that uncertainty is an important environmental variable in understanding switch costs. The authors did not find a difference in switch costs across age groups. While these authors did not uncover the age-related differences they hypothesized, they did not suggest that individual differences in switching costs for difference age groups do not exist. Rather, they suggest that their manipulations created a situation that was so difficult that all age groups were effected. The task uncertainty in this setting was too demanding. This idea underscores the importance of understanding the impact of uncertainty in a task-switching environment. This task characteristic may be critical to understanding why performance can suffer in multitasking environments in organizations.

These age-related multitasking studies send some messages to selection practitioners wishing to predict performance in a multitasking environment. First, these studies show that individual differences can predict performance in demanding multitasking environments. These studies also illustrate that even when individual differences in performance are present, task characteristics still impact performance. In the next two sections research is presented that addresses the generalizability of the performance decrements of various task characteristics to applied settings.

Multitasking Research on Driving Performance

Perhaps the most prolific application of multitasking research has been on investigating the detrimental effects of secondary or simultaneous tasks (e.g., monitoring controls, talking on phones, etc.) on driving performance. While this research is often conducted in the lab, it is worth reviewing due to the orientation towards the application of multitasking findings.

In one of the earliest studies on the topic of multitasking performance applied to driving, Kahneman, Ben-Ishai, and Lotan (1973) looked at driving performance using a dual-task paradigm. These authors used a dichotic listening task created for a different study by Gopher and Kahneman (1971) which will be reviewed later. In this study, Kahneman et al. (1973) hypothesized that performance in a dual task situation would predict accident frequency in a sample of bus drivers. The authors found that individual differences in the dual task performance measure correlated with three different accident criteria at $r = .29$, $r = .31$, and $r = .37$. Thus, the authors concluded that there are meaningful individual differences in dual task performance and these differences can be used to predict important performance criteria. This study is interesting and different from most of the multitasking research, as is a lot of the work by Gopher and Kahneman, in that these researchers treated multitasking performance as a predictor of criteria of interest. These studies by Kahneman and Gopher are groundbreaking in that they are the first application of multitasking research in an applied setting that can reasonably be thought of as parallel to an employee selection situation where a measure of performance in multitasking environment is used to predict future performance.

In a series of recent studies (Strayer & Johnston, 2001; Strayer, Drews, & Johnston, 2003; Recarte & Nunes, 2003), researchers have begun applying multitasking performance findings to cell phone usage while driving. Strayer and Johnston (2001) found that the secondary task of talking on a cell phone decreased performance in the primary task of driving. Strayer and Johnston used a low fidelity proxy for driving, but they later replicated their findings with a high fidelity driving simulation (Strayer, Drews, and Johnston, 2003).

In this replication, the authors predicted that the well-learned and practiced task of driving would not be attended to while talking on the cell phone. In other words, Strayer et al. (2003) posited that engaging in the secondary task (cell phone usage) would lead to a performance decrement in the primary task (driving performance). While these tasks are much more applicable to real world situations, the study design was consistent with the traditional cognitive approaches to studying multitasking performance. The researchers varied the task characteristics of the driving task to test the impact of cell phone usage on driving performance.

Strayer et al. (2003) found that regardless of traffic density, means of driving performance (in terms of accident frequency, brake on-set, brake off-set, time to accelerate, and following distance) were consistently worse in the dual task condition relative to the single task trial. Further, the authors found that when participants talked on the cell phone while driving, their recognition of previously seen billboards was lower than when the participants did not talk on the cell phone.

The researchers concluded that talking on hands-free cell phones altered how drivers attend to the stimuli in their environment. Strayer et al. (2003) suggested that the dual task cost on performance associated with cell phone usage while driving manifests itself as "inattentiveness blindness." The lack of attentional resource spent on driving in the presence of the secondary task can lead to very serious safety consequences.

In another replication and extension of multitasking research, Recarte and Nunes (2003) researched the impact of mental activity on visual-attention and response-selection capacities while driving. Further, they tested whether or not the distraction of mental activity would impact the entire visual field as opposed to a limited effect on peripheral information. To test these ideas, the researchers designed a driving simulation where participants were placed in a situation where they had to perform three types of tasks. The participants' attention was divided between driving, visual target discrimination (presented various places in the visual field), and mental tasks involving resource assignment or time sharing. Specifically, these mental tasks involved responding to information delivered either live or on a cell phone. For example, participants in one condition were asked to give detailed accounts of where they were and what they were doing at a specific date and time.

Consistent with the findings of Strayer, Drews, and Johnston (2003) and Strayer and Johnston (2001), Recarte and Nunes (2003) found that when drivers perform mental tasks, for example, talking on the cell phone, the percentage of detected targets in the visual field reduced significantly. The authors note that this multiple processing cost could lead to traffic accidents. The authors did not find a "tunnel vision" effect on the visual field. Rather, they found that the introduction of a mental task in the driving situation impacted the entire visual field.

The importance of these driving performance studies lies in the setting in which they took place. The research discussed above investigated how laboratory-based task characteristic multitasking findings extend to applied environments with a high demand to multitask. Where the age-related individual difference studies discussed earlier establish that individual characteristics can predict performance in a multitasking environment, these driving studies show that task characteristics are important performance antecedents in applied settings, as well as the lab. While neither of these streams of research are particularly job related, they do show organizations that lab research on tasks characteristics may generalize to applied organizational settings, and that measuring individual differences helps organizations understand and predict performance in job-related multitasking environments. In the next section a series of studies is discussed where multitasking task characteristics research is applied to a job-relevant situation and individual differences in the multitasking situation are used to predict the future job performance of airplane pilots.

Multitasking Research on Pilot Performance

As mentioned earlier, driving performance is not the only arena where multitasking research has been applied. Another area of application from this stream of cognitive research has been on the performance of airplane pilots. Like the Kahneman et al (1973) study, these studies are significant because they treat multitasking as an ability that predicts job performance outcome measures. This line of research could prove to be the first step along the way for selection practitioners wishing to use multitasking measures as a selection component.

Gopher and Kahneman (1971) conducted a study that highlighted the importance of multitasking for pilots 35 years ago. These authors point out that pilots-in-training often fail due to their inability to divide attention among concurrent activities or they are slow to do so. Thus, the researchers conducted a study to identify key differences in multitasking performance that predict actual job performance outcomes.

The authors sampled pilots-in-training in the Israeli Air Force. Two messages were simultaneously delivered to the participants in opposite ears. The researchers hypothesized that individual differences in performance on the dichotic listening task would predict success as a pilot.

Gopher and Kahneman (1971) found that errors in the dichotic listening task correlated with a flight school criterion measure with magnitudes ranging from $r = .10$ to $r = .36$. Interestingly, although the authors did not discuss this finding, they reported in a table of correlation coefficients that intelligence related to dichotic listening performance at r = .21, which suggests that individual differences in intelligence may explain variance in multitasking performance.

Eleven years after the Gopher and Kahneman (1971) study, Gopher (1982) replicated the earlier findings. In this study, Gopher reported the results of a criterion-related validity study where a dichotic listening task was used to select pilots as part of a battery of selection tests given to candidates for the Israeli Air Force. This study is one of the very few efforts to understand the role of multitasking performance in the context of employee selection.

To understand the validity of the inferences drawn from the same dichotic listening task that Gopher and Kahneman (1971) employed, Gopher (1982) looked at results of a sample of 2000 flight training candidates. The author noted that suitable criterion measures were difficult to collect and thus they used more or less a training pass/fail measure of success on the job. Gopher (1982) searched for mean differences in dual processing performance in the dichotic listening task across unsuccessful and successful flight training participants. The researcher found large and significant mean differences across groups for all of the performance measures in the dichotic listening task. A more general conclusion reached by Gopher was that the results show individual differences in dual task performance do exist and that these individual differences correlate with external criteria. This conclusion suggests that looking for characteristics that can predict performance in a job relevant multitasking environment could be useful in generating information about how best to predict job performance in high multitasking demand jobs.

More recently, O'Hare (1997) conducted yet another criterion-related validity study on a common battery of tests used to select pilots. While the tasks completed by the participants in this project were not given a simple dichotic listening task like Gopher (1982) and Gopher and Kahneman (1971), they did have to perform multiple tasks in the same general time period using a custom-designed 13-button keyboard and joystick. The participants had to complete several tracking tasks. The participants received ample practice on these tasks. This test measures situational awareness when confronted with an environment that requires multiple task switches. To validate the inferences drawn from this test the authors sampled a group of pilots and a group of non-pilots to look for performance differences. The authors found significant mean differences in the two groups on the multitasking measure. It should be mentioned, however, that this study did have some serious weaknesses. The chosen criterion measure (group membership as a pilot) is not a particularly meaningful and the sample size in both groups was under 20.

Through this research on age, driving performance, and pilot performance, the lab research on multitasking performance is beginning to get applied. The age research has shown that meaningful individual differences in the ability to perform multiple tasks in the same general time period exist. The driving performance research has shown that task characteristics similar to those varied in lab experiments can lead to important performance deficiencies in applied settings too. The pilot research has shown measuring individual differences on dual task performance and task-switching performance can predict success on various job-related criteria.

This collection of findings highlights the importance of understanding performance in multitasking environments in the work place. This applied research suggests that: (a) we can find individual difference predictors that explain job relevant multitasking performance variance (Linkage 2 in Figure 1); (b) task characteristics will impact performance in a job-related multitasking environment (Linkage 1 in Figure 1); and (c) understanding how task characteristics and individual differences impact performance in a multitasking environment will help organizations predict who can survive in multitasking environments. As pointed out by Zaldivar (2006), at this point these ideas have gone relatively unexplored in organizational settings. The few studies conducted on the topic are reviewed in the next section.

Multitasking Research Applied to Organizational Contexts

The research practice gap between applied multitasking research and the need to understand performance in multitasking work environments is starting to receive some attention, but not much has developed yet. A review of typical I/O Psychological research outlets turns up very little on the topic. What has been done so far is thought provoking and highlights the need for further inquiry. These few attempts to understand performance in multitasking environments in the workplace are reviewed in this section.

Perhaps the most comprehensive attempt at identifying meaningful and applicable individual differences that predict performance in multitasking environments to date was an unpublished doctoral dissertation (Delbridge, 2000). This work attempted to comprehensively describe and test a nomological network of “multitasking ability” targeting cognitive abilities and personality traits antecedent to differences in performance in a multitasking situation.

Delbridge (2000) suggested that environments high in the need for multitasking have several key characteristics. These characteristics include a frequent task switching, uncertainty about when task switches will be needed, and time pressures to complete tasks. The author's theory was that people differ in their ability to perform in these situations. These performance decrements are manifested through process losses due to uncertainty; reaction time losses; and stress, coping, and withdrawal. The author also suggested that many individual difference variables predict performance in such a multitasking environment. These differences in individual attributes impact performance due to their impact on the different losses associated with the multitasking environment. For example, if stress tolerance can predict coping strategies, then stress tolerance will also in explain variance in performance in a multitasking environment.

Beyond the stress tolerance example, the author proposed several other individual difference predictors as well. Delbridge (2000) predicted that intelligence, tolerance for ambiguity, openness to experience, neuroticism, coping styles, performance orientation, mastery orientation, trait anxiety, and conscientiousness explain performance variance in multitasking environments.

Delbridge (2000) tested this proposed nomological network by collecting data from 232 psychology students. The individual difference data was collected using commonly accepted tests and surveys (e.g., the Raven's Progressive Matrices test was used to measure cognitive ability). The multitasking performance measure was created for the purposes of the dissertation. In the simulation, the author created a scenario consisting of three simple tasks. Unlike the traditional cognitive lab research, these tasks were designed to emulate real job performance, although there is some concern about whether or not this goal was achieved. The participants had to look up numbers or other information in a chart, respond to situational judgment test items designed to tap managerial potential, and respond to logic puzzles containing job-related content. Performance was defined in terms of accuracy on these tasks.

Delbridge (2000) found mixed results for her hypotheses. She found that the multitasking environment did result in the predicted performance cost. However, the author was unable to find any of the predicted influences of individual difference variables on performance in a multitasking environment. The author did note, however, that the data addressing many of the hypotheses were trending towards the predicted direction.

One possible explanation for the lack of significance of these finding could be low statistical power. The author's hypotheses were by and large mediation predictions. These interaction hypotheses probably could have been more easily identified if the author had a larger sample size. Further, the author reported that the means were quite low on some of their tasks and that there was not a lot of variance. This lack of variance in the criterion measure could have further magnified the power problem in the study. Between the statistical power problem and the problems observed in the criterion measures, the lack of support for Delbridge's (2000) hypotheses should not be considered conclusive; further research is warranted around her predictions.

Methodological and measurement issues aside, the Delbridge (2000) dissertation accomplishes several steps in the direction of applying lab multitasking research to work environments. This study represents one of the few scientific resources to date that explicitly discusses the role of multitasking in organizations. Further, this dissertation outlines a cohesive argument for several individual difference variables that should be considered relevant to understanding performance differences in work-related multitasking environments. The arguments highlighted by Delbridge (2000) may prove to be a pivotal conceptual development in this research area.

Ishizaka, Marshall, and Conte took the understanding of the role of individual differences important to multitasking performance a step further in a 2001 Human Performance article. These authors were interested in understanding how Type A Behavior Patterns (TABP) predicted performance in a multitasking environment. To test this hypothesis, the authors collected data from 118 psychology students. The participants first completed the predictor survey. Then they completed a multitasking simulation. This simulation consisted of two visual tasks involving the completion of simple math problems. These problems were completed while simultaneously monitoring and controlling the levels of six different gauges presented on a different part of the screen. The participants also had to complete an auditory task that involved attending to and encoding words presented out loud during the eight minute session. The participants were asked to recall the words at the end of the trial.

The authors found several interesting relationships between performance in the multitasking simulation and the individual differences that were measured. First, the authors did not find that global TABP predicted performance on the multitasking criterion measure. They did find significant relationships between performance and facets of TABP, such as scheduling, time urgency, list making, and achievement striving. The authors suggested that these relationships can be explained by the differences in attentional resource allocation between high and low TABP people. That is, high TABP people attend to different tasks and different aspects of performance in a multitasking environment than do low TABP people.

While this study was not conducted in an organizational setting and thus has some generalizability questions, the work of Ishizaka et al (2001) is significant in a few ways. First, the nature of the task in this simulation was non-trivial and work-relevant in comparison to much of the cognitive multitasking research. Second, unlike the age research reviewed earlier, this study addressed individual difference variables highly relevant to work settings. Ishizaka et al (2001) presented the first study to investigate critical organization-relevant individual difference variables in a job-related multitasking environment. This study is the first real confirmation that researching such relationships can provide meaningful results.

Konig, Buhner, and Murling extended this individual difference predictor research further in another Human Performance article in 2005. These authors investigated the role of working memory, fluid intelligence, attention, polychronicity, and extraversion in predicting performance in a multitasking environment.

These researchers tested the impact of these individual difference variables in a sample of 132 undergraduate and graduate students in a German university. To measure multitasking, Konig et al (2005) used the SIMKAP (Bratfisch & Hagman, 2003; Bratfisch, Hagman, & Puhr, 2002), which is a complex standardized multitasking simulation designed for personnel selection purposes. This multitasking simulation consists of list identification tasks and simple logical-verbal, logical-numerical, and arithmetic tasks. While the participants complete these tasks, they have to periodically switch to scheduling tasks; where they have to respond by looking up information on a computerized calendar. Performance is defined by an aggregation of reaction time and accuracy measurements. The developers of the tool reported reliability coefficients in the .90's (Bratfisch & Hagman, 2003). The complexity of the tasks in this simulation is more parallel to the work environment than any other measure of multitasking used to date, although Russell and Sacco (2006) suggested that the SIMKAP may have some face validity concerns.

The predictors were measured using a series of sensible tools. Attention was measured using the Test Battery for Attentional Performance (TAP; Zimmerman & Fimm, 1993). Working memory was measured using three tests: Reading span, spatial coordination, and switching numerical. Fluid intelligence was measured using the Intelligenz-Struktur-Test 2000R, which is a well established German intelligence test (Schmidt-Atzert & Deter, 1993). Polychronicity was measured using the individual version of the Inventory of Polychronic Values (Conte & Jacobs, 2003; Conte, Rizzuto, & Steiner, 1999), which was adapted from a scale created by Bluedorn, Kalliath, Strube, and Martin (1999). Finally, extraversion was measured using a 12-item German extraversion scale created by Borkenau and Ostendorf (1991; 1993).

Using confirmatory factor analysis, Konig et al (2005) found that the SIMKAP data loaded onto three factors, two of which were accuracy measures and one was a reaction time measure. The individual difference variables predicted SIMKAP performance nicely in a series of hierarchical regression analyses. Overall R-squared values ranging from .12 to .27 were observed across the three factors falling out of the SIMKAP confirmatory factor analysis. Most of this variance was explained by intelligence and working memory. Attention also explained significant levels of performance variance. Polychronicity and extraversion added very little explained variance to the model regardless of the order of the steps.

This study is the most comprehensive study of individual differences in a multitasking environment to date. Konig et al (2005) found several important predictors of performance in a job-relevant (in terms of task complexity) multitasking simulation. There remain some generalizability to organizational settings concerns due to the use of a student sample, but this study is an excellent example of the potential for individual difference variables to explain job performance in multitasking environments.

While progress has been made, with few exceptions predicting performance in multitasking environments has been difficult in organization-relevant applied settings. Given that mixed results have been observed even in recent research in controlled lab settings (Lien, Proctor, & Ruthruff, 2003; Tombu & Jolicoeur, 2004), perhaps the difficulty lies, in part, in a criterion problem. As discussed in the review of the Delbridge (2000) study, in many applied studies of performance in a multitasking environment the tools used to measure performance leave something to be desired. As the Konig et al (2005) study illustrates, if care is taken to develop the criterion measure so that it parallels the key features of a job-relevant multitasking environment, then meaningful individual difference predictors of performance in that situation emerge. Therefore, understanding the key features of a multitasking environment is critical to choosing or developing a criterion measure for use in understanding the role of individual difference variables in performance in these complex environments (Linkage 2 in Figure 1).

Key Characteristics of a Job-Relevant Multitasking Environment: Understanding Why Performance Varies in these Situations.

Some common themes can be identified in the multitasking research reviewed up to this point. Applied multitasking environments are those where there is a frequent need to switch between and reprioritize tasks, there is no perceivable pattern or predictable due to switch tasks, and there is a salient sense of urgency or time pressure (Delbridge, 2000). These characteristics are the critical features of multitasking environments and performance in these situations depends on how well an individual can manage these three situational demands.

Any multitasking environment must involve switches from one task to another. Something in the situation must indicate when the next task is available to perform. In order to switch tasks, an individual must register the cue that directs the switch, stop performance on the current task, and commence performance on the next task. This switch can be costly in terms of performance even when the two tasks are simple and similar in nature (Gopher & Kahneman, 1971). Variance between people in their ability to switch tasks can explain important performance differences in multitasking environments. The likely antecedents of this variance, as suggested by years of cognitive multitasking research, are differences in "inspection time" and "reaction time".

Inspection time refers to how long a cue to switch tasks is in an environment before it registers with the individual. An example of measuring inspection time comes from Deary and Stough (1996). These researchers presented visual stimuli to participants and followed the stimuli with simple questions about what the individuals observed. The questions were designed to be easy enough for any person with reasonable ability to answer the questions correctly if given enough time to study the image. The measure of inspection time was a record of the shortest times the individual needed to correctly answer the question. In a multitasking environment, individuals who attend to the cues to switch tasks more quickly will be more effective at switching tasks.

After the need to switch tasks is registered, then reaction time is an important factor impacting how the switch occurs. Those that can react quickly should be able to switch tasks quickly. Salthouse et al. (1998) found a significant relationship with reaction time and performance deficiencies due to task switching.

Another key feature of a multitasking environment is uncertainty (Delbridge, 2000). In other words, in applied multitasking environments, individuals cannot predict when the need to switch tasks will arise. Anderson and Kida (1985) pointed out that this unpredictability changes people's performance-related attributes. Applied to an organizational context, such as personnel selection, this finding suggests that if some selection component is able to predict task performance in a single task environment, the predictive validity of that tool may change in a situation characterized by a need to switch tasks at uncertain intervals.

In support of this idea, De Jong (1995) found that individuals were able to switch tasks faster and their performance was more accurate after the switch if they knew when the indicator to switch tasks would arrive. The lesson from this research is that unpredictability in task switching may be a key component of the environment that can lead to performance differences in high multitasking jobs.

The final component of a multitasking situation is time pressure or urgency. Without a sense of urgency, individuals can complete tasks sequentially without having to divide attention between the present task and the cue for the next task. Reprioritization of tasks likely never happens. In lab settings, time pressure is not as critical because researchers define for participants when to switch tasks, but in organizational settings time pressure is what drives the need to switch tasks.

Delbridge (2000) suggests that the mechanism through which time pressures effect performance is the stress cycle. She suggests that individual differences in the interpretation of time pressure as a stressor and the resulting coping strategies are the means through which situations with high degrees of time pressure impact performance. This assertion has been supported in research literature on stress (e.g., Mann, 1992). Further, Nygren (1997) found that stress changes the way decisions are made. This finding suggests that if time pressure kicks-off the stress cycle in multitasking environments, decisions about the prioritization of tasks could be impacted. Thus, performance in multitasking environments may vary, in part, on how individuals perceive time pressures and how individuals handle stress (Delbridge, 2000). For example, if a person tends to withdrawal or become frustrated in the face of a time pressure stressor, they may be less likely to excel in that situation than a person who takes a problem-focused coping approach. If these time pressures exist, variance in performance should be expected and the stress cycle maybe the mechanism through which this variance is created.

Understanding these key components of an applied multitasking environment can provide critical insights into how best to pursue the needed link between science and practice in the area of multitasking research and performance in demanding multitasking environments in organizations. While much of the multitasking research has not offered, for example, specific guidance about how to select employees into a high multitasking setting, it has identified the key components of a multitasking situation and given clues as to how and why performance in these situations may vary.

Armed with this understanding of the key features of multitasking environments, researchers can begin to investigate a theory-driven approach to predicting performance in jobs with high multitasking demands. In order to so, research must 1) identify meaningful individual differences that predict performance in a job relevant environment with these key environmental components present (Linkage 2 in Figure 1), and 2) understand how the task characteristics explored in laboratory multitasking research apply to job relevant environments with these key environmental components present (Linkage 1 in Figure 1). The focus of this study is to investigate various task characteristics (see Figure 2) and individual differences (see Figure 3) that lead to varying performance levels in a job relevant multitasking simulation. The simulation was developed to measure job candidate performance in a high fidelity, realistic, job-related, environment with the key components of a demanding multitasking situation present. Details of the development of this tool are provided in the Appendix and Table 1. A screenshot of the multitasking simulation is provided in Figure 4. In the next sections, hypotheses predicting task characteristics and individual difference effects on performance in a multitasking environment are explained.

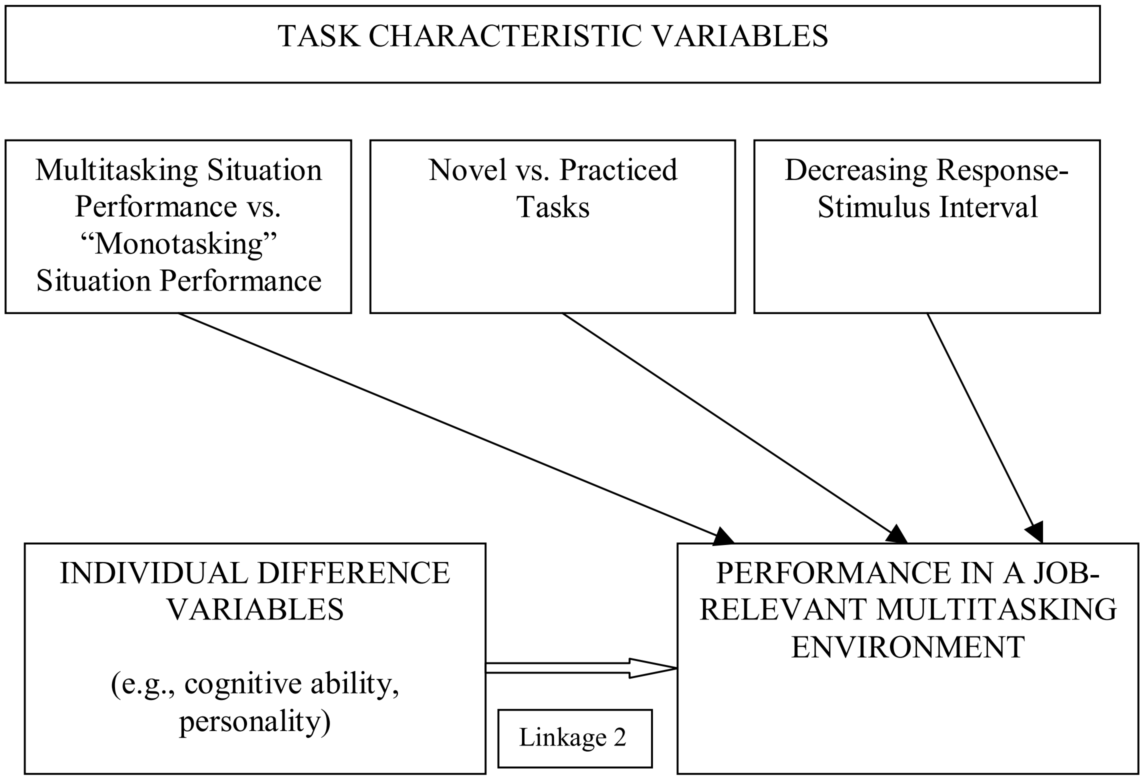

Figure 2. Conceptual Model of Task Characteristics under Study in the Thesis.

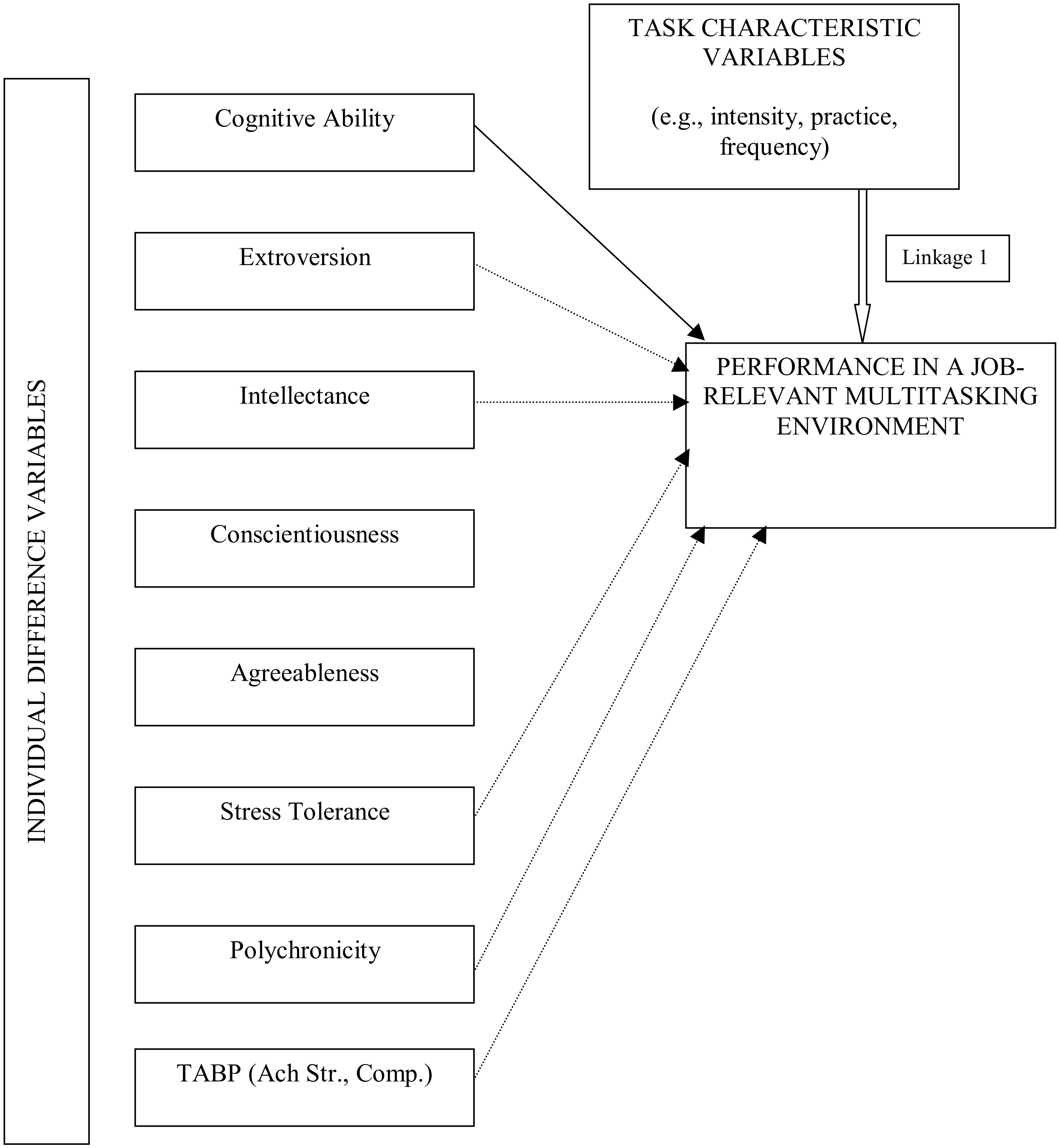

Figure 3. Conceptual Model of Individual Differences under Study in the Thesis

Table 1

Descriptive Statistics Pattern from the Initial Concurrent Validation Study (Incumbent Sample n = 226).

Trial #	Query #	Delivery Time	Min.	Max.	Mean	SD
Trial 1	Query 1	Start	.00	1.00	.71	.33
	Query 2	55 sec. Later	.00	1.00	.68	.31
	Query 3	50 sec. Later	.00	1.00	.65	.30
	Query 4	45 sec. Later	.00	1.00	.68	.35
	Query 5	40 sec. Later	.00	1.00	.59	.41
	Query 6	35 sec. Later	.00	1.00	.43	.35
	Query 7	40 sec. Later	.00	1.00	.48	.44
	Query 8	40 sec. Later	.00	1.00	.56	.33
Trial 2	Query 1	Start	.00	1.00	.77	.29
	Query 2	55 sec. Later	.00	1.00	.80	.29
	Query 3	50 sec. Later	.00	1.00	.83	.27
	Query 4	45 sec. Later	.00	1.00	.80	.31
	Query 5	40 sec. Later	.00	1.00	.60	.41
	Query 6	35 sec. Later	.00	1.00	.56	.42
	Query 7	40 sec. Later	.00	1.00	.63	.43
	Query 8	40 sec. Later	.00	1.00	.66	.33

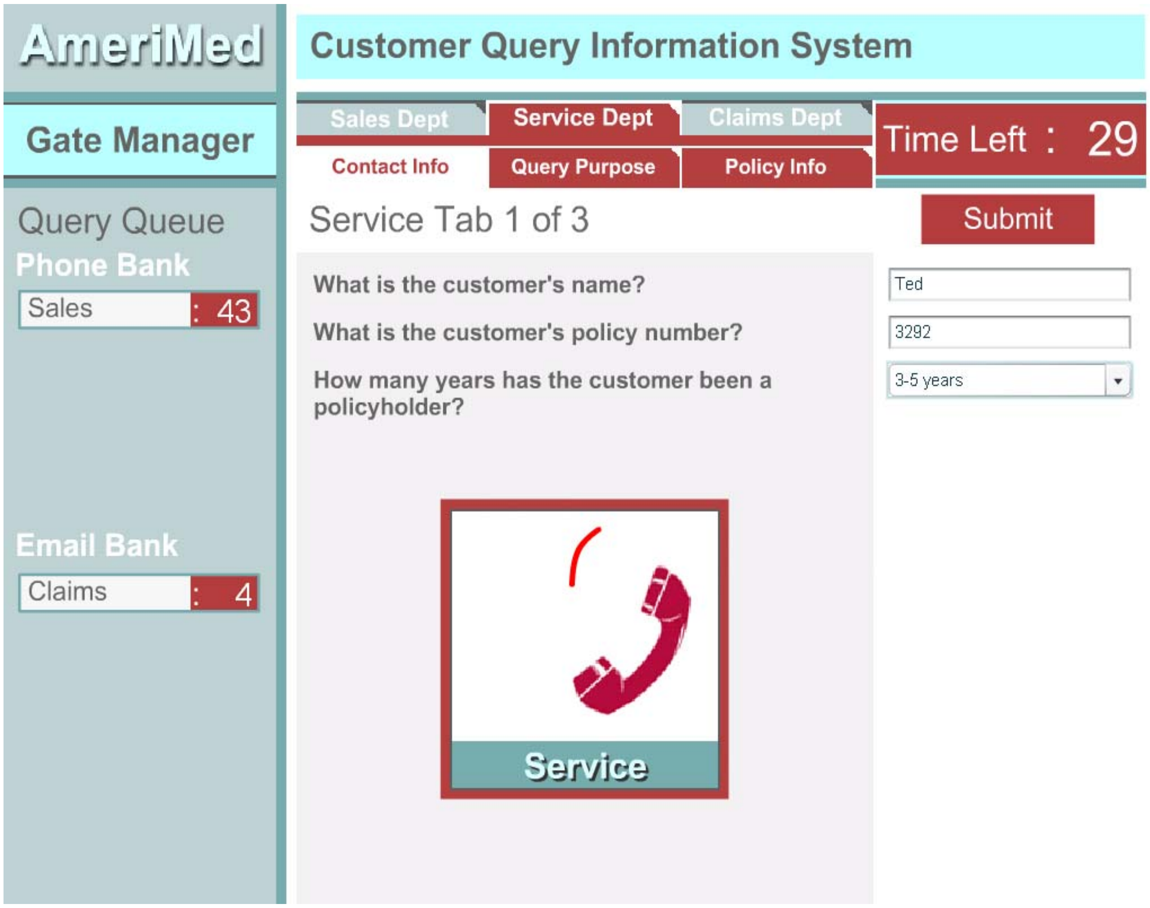

Figure 4. A Screenshot from the Multitasking Simulation.

Hypotheses about Performance in a Job-Relevant Multitasking Environment

The literature reviewed above identifies several streams of multitasking-relevant research that would benefit from further exploration. Using the multitasking simulation explained in the Appendix, data can be evaluated to test two important types of hypotheses. First, as pointed out in the literature review, there has been a lot of research on the human capacity to multitask across varying task characteristic conditions and this research has furthered our understanding of human information processing. There have not been many replicating or extending studies investigating these findings in applied settings. Given that several of the manipulations common to lab-based multitasking research were intentionally built into this tool, this simulation provides a unique opportunity to test the generalizability of several multitasking research laboratory findings (Linkage 1 in Figure 1) in applied settings.

Second, the literature review points out that not much research has accumulated about how individual characteristics lead to performance differences in on-the-job multitasking environments. Despite this lack of research, several streams of research have investigated how performance is impacted by the situational characteristics discussed in the last section (the key characteristics of a multitasking environment). For example, research has suggested that cognitive ability predicts task switching performance and aspects of personality may predict performance under time pressures. While this research typically does not address multitasking at work directly, it can be leveraged to form a series of hypotheses about how cognitive and non-cognitive predictors relate to performance in a multitasking environment. Thus, several individual difference attributes will be discussed in terms of how they are likely to vary across different multitasking environmental components and hypotheses will be tested to further our understanding about individual difference predictors of performance in a high fidelity applied multitasking environment (Linkage 2 in Figure 1).

Task Characteristic Hypotheses

Several replications and extensions of the cognitive multitasking research are possible using the data collected with the multitasking simulation. As was discussed in previous sections, the bulk of the body of our understanding in the human capacity to multitask comes out of lab studies observing trivial tasks (e.g., Yeung & Monsell, 1995;

Klapp et al, 1991, Meiren, 1996; Gopher et al, 2000, etc.). In the present study, the simulation used to collect criterion data is consistent with many of the experimental conditions tested in these studies with one important difference. The multitasking simulation consists of applied, job-relevant, non-trivial tasks. Bridging this gap between the lab and applied settings could be an important step forward in our understanding of the impact and implications of multitasking in the workplace.

One such finding that can be examined using the multitasking simulation is that performance suffers as the demand on attentional resources increases. This is the basic finding of most lab dual task and task switching research and has been observed in study after study dating back to Gopher and Kahneman (1971). The idea is that performance on one task will suffer if multiple tasks enter the environment to which attention must be allocated. This hypothesis is typically tested by allowing candidates to practice both tasks in a single task environment, then the tasks are put together, and the performance decrement is measured (e.g., Lindenberger et al, 2000). This methodology effectively tests the hypothesis, but this is not how multitasking environments exist in applied settings. Rather, in the workplace the primary task is interrupted with unknown tasks and at uncertain intervals, but the performance decrement theoretically should still exist because key characteristics of a multitasking environment are still present. The multitasking simulation affords the opportunity to test whether or not the typical dual task performance decrement replicates to an applied setting.

Hypothesis 1 – An adverse effect on performance relative to a single task environment will occur when attentional resources must be shared across multiple tasks.

Another research question that has received attention regards the impact of practice on performance in multitasking situations (Yeung & Monsell, 2003; Van Selst, 1999; Gopher et al, 2000; Hazeltine et al, 2002). The findings from this research have been mixed. Yeung and Monsell (2003) found that the multitasking performance decrement was robust despite the opportunity to practice. Hazeltine et al (2002) were able to eliminate the impact of the dual task situation through practice. Van Selst (1999) observed the dual task performance decrement but noted that it improved with practice. The lab research is mixed and these differences are probably driven by the characteristics of the task. The tasks employed in these lab studies are usually trivial, but novel. Due to the novelty, neither task is automatic or overlearned, thus attention must be spent on both tasks and a decrease in performance is observed, which is consistent with Yeung and Monsell (2003). That said, given the nature of the tasks, if one had an infinite amount of time to over learn the tasks, the performance decrement could likely be eliminated, such as the case with Hazeltine et al (2002). In this study, the researchers had the participants repeat trials until the cost on performance was eliminated. Neither of these situations occur in applied settings, however. In most applied settings, people complete tasks that are not entirely new to them, but they only get one chance to do it. For example, a call center employee only gets one chance to sell to any individual customer, but the task of talking on the phone occurs over and over again in the course of a shift. Thus, in applied settings, results consistent with Van Selst (1999) are likely to be observed.

Hypothesis 2 – Practice on the simulation tasks will reduce, but not eliminate the performance decrement created by the multitasking environment

In an attempt to understand the task-switch adverse performance effect commonly found in the lab studies, Rogers and Monsell (1995) conducted experiments to investigate what task characteristics in the multitasking environment lead to reduced performance. These authors found support for the hypothesis that the performance decrement is tied directly to the response-stimulus interval. That is, performance decreases as the time difference between the response on task one and the presentation of the stimulus for task two decreases. The response-stimulus interval is one of the key manipulations in the multitasking simulation. With each successive task, the interval decreases. Thus, the multitasking simulation is an ideal setting in which to replicate and extend Rogers and Monsell's (1995) findings to an applied setting.

Hypothesis 3 – As the time interval between the response to the previous task and the delivery of the next task decreases, task performance will also decrease.

The task characteristic hypotheses are listed in Table 2.

Table 2

Proposed Task Characteristic Hypotheses

#	Hypothesis
1	An adverse effect on performance relative to a single task environment will occur when attentional resources must be shared across multiple tasks.
2	Practice on the simulation tasks will reduce, but not eliminate the performance decrement created by the multitasking environment
3	As the time interval between the response to the previous task and the delivery of the next task decreases, task performance will also decrease.

Individual Difference Hypotheses

While the task characteristic hypotheses discussed above are likely to shed light on the dynamics of performance in a job-related multitasking environment, perhaps of more interest from a traditional Industrial/Organizational Psychological perspective is understanding individual differences that can explain variance in these settings. Gaining an understanding of how different individual difference variables predict performance in demanding multitasking jobs would be tremendously useful to organizations with a need to select people into such environments. In the next sections various individual difference variables are discussed relative to multitasking performance and hypotheses are suggested about how these variables will predict performance in a job relevant multitasking environment (Linkage 2 of Figure 1).

Cognitive Ability

As Ree and Carretta (2002) point out, cognitive ability has been the most robust and consistent predictor of job performance that has emerged in 100 years of psychological research. Several meta-analyses based on hundreds and hundreds of studies have identified how well this ability predicts performance across time, industry, culture, job complexity, etc. (e.g., Schmitt, Gooding, Noe, & Kirsch, 1984; Hunter & Hunter, 1984; Murphy, 1989; Barrick, Mount, and Strauss, 1994; Schmidt & Hunter, 1998). In fact, cognitive ability has received so much attention that Murphy (1996) stated that I/O research can be separated in to two types: 1) Studies investigating the relation between cognitive ability and job performance, and 2) everything else, with type 1 receiving the bulk of the attention.

This broad tradition of cognitive ability research has led to a very clear conclusion. Cognitive ability is the single best individual difference variable available when it comes to predicting job performance, regardless of the circumstances (Ree & Carretta, 2002; Gottfredson, 2002; Reeves & Hakel, 2002; Murphy, 2002, Schmidt, 2002; Goldstein, Zedeck, & Goldstein, 2002). Gottfredson (1997) stated that "no other measurement trait, except perhaps conscientiousness, has such general utility across the sweep of jobs in the American economy" (pg. 83). With such a rich history of predicting job performance, cognitive ability is the logical starting point in thinking about abilities and attributes that can predict job performance under a high multitasking demand.

Research on cognitive ability has uncovered that intelligence can be adequately represented by a hierarchy of abilities with a single factor residing at the upper most level (Carroll, 1993). This single factor, often referred to as 'g', represents the common core present in all tests of cognitive ability (Spearman 1904). Over the years, findings have shown that all tests of intelligence load onto this common latent factor. Regardless of the content (e.g., spatial, verbal, math) and methods (computer-based, psychomotor, oral, paper and pencil, etc.) tests of cognitive ability all tap the same general mental ability proposed by Spearman in the early 1900's (Ree & Carretta, 2002; Gottfredson, 2002).

This hierarchical model of cognitive ability was fully explained by Carroll's (1993) three-stratum model of cognitive ability. This model is based largely on 100 years of psychometric research on the topic. The topmost level in the hierarchy is a single 'g' factor. Gottfredson (2002) described 'g' as the ability to learn moderately complex material quickly and avoid cognitive errors; she suggests that it is the ability to process information.

At the next step down in the hierarchy, there are a number of broad abilities including fluid intelligence, crystallized intelligence, general memory ability, broad visual perception, broad auditory perception, broad retrieval ability, and broad cognitive speediness. While rank orders of scores on these abilities will not correspond exactly, the presence of the general 'g' factor suggests there is significant overlap (Murphy, 2002).

Finally, in Carroll's 1993 model, these broad abilities each have sub-levels consisting of more specific abilities. Murphy (2002) provided examples of these specific abilities for each of the broad-stratum abilities: induction – fluid intelligence; language development – crystallized intelligence; memory span – general memory ability; spatial relations – broad visual perception; sound discrimination – broad auditory perception; word fluency – broad retrieval ability; and perceptual speed – broad cognitive speediness. Murphy (2002) said that the three-stratum model suggests that general measures of 'g' will predict performance due to the overlap among the sub-stratum, even if the task requires a specific sub-ability. For example, if performance in a multitasking environment requires a specific cognitive ability, such as memory capacity, any construct-valid measure of cognitive ability should predict performance in that environment.

While this hierarchical model approach to understanding cognitive ability is useful, another approach to cognitive ability has emerged out of cognitive psychology over the last couple decades (Ben-Shakhar & Sheffer, 2001). The cognitive psychology approach to researching intelligence has focused on understanding how information-processing resources relate to individual differences in cognitive ability. Bazana and Stelmack (2002) point out that the last 20 years has seen progress in our understanding of individual differences in mental ability through the application of cognitive information processing models. These authors cited several relevant examples, such as, the observed relationships between psychometric ‘g’ and response times on sensory, motor, memory, and decision tasks; and the speed of information processing.

This cognitive psychological approach supplements our understanding of cognitive ability, but the information processing models of intelligence do not seem to measure a distinctly different construct from psychometric ‘g’ (Stauffer, Ree, & Carretta, 1996). This statement is not universally endorsed (e.g., Stankov, 1988; Kyllonen & Christal, 1990), and there still remain questions about the incremental validity of cognitive components measures beyond traditional psychometric measures of cognitive ability (e.g., Roznowski, Dickter, Hong, Sawin, & Shute, 2002). That said, there is a sound theoretical underpinning and consistent pattern of results linking individual differences in information processing to ‘g’.

Even among those that disagree, it is acknowledged that there is a strong relation between the information processing model of intelligence and the hierarchical model of cognitive ability. As Gottfredson (2002) pointed out, tasks that impact the information processing abilities place a premium on ‘g’. She further suggests that along these lines, multitasking and dual processing tasks also put a premium on ‘g’. Further, research has shown that reaction time (Salthouse et al, 1998) and inspection time (Deary and Stough, 1996) yield a strong information processing factor coinciding with ‘g’. This finding even extends to the physiological manifestations of the information processing system. Reed and Jenson (1992) observed a correlation between intelligence and the physiological speed of neural processing ($r = .37$). Nettlebeck and Lally (1976) suggested that speed and scanning mechanism differences in information processing lead to differences in intellectual ability. Lastly, Cornelius, Willis, Nesselroade, and Baltes (1983) and Hunt (1978) stated that differences observed in indices of intelligence may reflect various individual differences in information processing, such as accessing overlearned code, attentional capacity, accessing and manipulating information in active memory, and resistance to interference or distraction.

Understanding the more recent cognitive psychology advances in our understanding of human intelligence in concert with the traditional hierarchical model clearly illustrates the commonality between cognitive ability and performance in a multitasking environment. As discussed earlier, multitasking environments are characterized by several key components (e.g., frequent task switches) that cognitive research has identified as being impacted by the human information processing systems. Thus, the bottom line is that if there is a link between multitasking performance and information processing, given that information processing loads onto a general 'g' factor as discussed above, then any construct valid measure of cognitive ability should predict performance in a multitasking situation.

Fortunately, there is a lot of research that has investigated how information processing influences performance in situations relevant to multitasking environments. Some of these studies are discussed below to illustrate the linkage between performance in multitasking situations, information processing and cognitive ability.

Ben-Shakhar and Sheffer (2001) investigated the amount of available resources and the ability to allocate them among various components of a task. They suggested that identifying individual differences in this multitasking situation would account for individual differences in problem solving and general intelligence. These authors used a dual task paradigm to investigate performance differences that could not be accounted for by single task performance and they investigated the relations between dual task performance and measures of general cognitive ability. Ben-Shakhar and Sheffer (2001) found that dual task performance was related to general ability and that this relationship was higher than the relation between single task performance and cognitive ability. They did, however, find that these effects dissipated with practice, but this may have been due to the simple nature of the tasks used in the study.

The relation these authors found between dual task performance and cognitive ability was consistent with results found by Stankov (1988; 1989). This author also suggested that 'g' predicts dual task performance above and beyond the single tasks comprising the dual task situation. To explain these results, Stankov (1988;1989) explained that higher ability individuals choose more effective cognitive strategies and learn more rapidly. Further, higher ability people may have more attentional resources (also suggested by Ben-Shakhar & Sheffer in 2001). These findings are consistent with the finding that task complexity mediates the relationship between cognitive ability and performance (Roberts, Beh, & Stankov, 1988; Schmidt & Hunter, 1998; Gottfredson, 2002; Tenopyr, 2002).

Stankov's (1988; 1989) studies have a couple important implications in linking cognitive ability to performance in a multitasking environment. First and foremost, these results show that performance in multitasking situations is influenced by cognitive ability. Second, these authors discuss the information processing features that could be linking multitasking performance to cognitive ability. Stankov's (1988; 1989) work explains that differences in cognitive ability influence how people register and encode stimuli in their environment. Based on these studies, higher ability people are able to notice, understand, devise a strategy, and allocate resources to the tasks in a multitasking environment more effectively than lower ability people.

Salthouse et al (1998) investigated the role of task switching as a moderator between age and cognitive ability. They administered tests of inductive reasoning, spatial visualization, and free recall to 161 participants. They found that task switching in terms of reaction time and accuracy impacted the relationship between age and cognitive ability tests. Correlation coefficients between the task switching variables and cognitive ability ranged from the mid .20 to nearly .50. They stated that "faster switchers have higher levels of cognitive performance" (pg. 460). This finding is of direct relevance to the present study because it shows that cognitive ability has a strong relationship with measures of task switching, which is one of the key components of the job-related multitasking simulation.

Also informative, there have been studies that have focused solely on the link between information processing and psychometric measures of intelligence. Cornelius et al (1983) administered four second-stratum tests of cognitive ability (fluid intelligence, crystallized intelligence, short-term acquisition and retrieval, and broad speediness). They also administered four tests of different aspects of attention to measure information processing (decoding processes, selective attention, attention switching, and concentration). They found significant, moderate, relations between the information processing measures and the psychometric cognitive ability measures. This finding highlights the relationship between several of the information processing components needed in a demanding multitasking environment and psychometric measures of cognitive ability.

Similarly to Cornelius et al (1983), Kyllonen and Christal (1990) investigated the relationship between information processing and aspects of psychometric cognitive ability. These authors hypothesized that while there is a general ability factor of intelligence from the traditional psychometric research, no such parallel construct of information processing exists. To test this hypothesis, the authors administered several measures of working memory across several of its different defining features (simultaneous processing and storage capacity, and numerical and linguistic processing capacity). Kyllonen and Christal (1990) found that all of the working memory features loaded onto a general working memory factor. They also found very strong relationships between the working memory factor and reasoning ability. In fact, they suggest that reasoning ability and the working memory factor are similar if not identical constructs. This study is relevant to the current research because it shows again that aspects of information processing important to performance in a multitasking environment are also highly related to cognitive ability.

Stauffer, Ree, and Carretta (1996) asked the same question as Kyllonen and Christal (1990), but from the opposite direction. While Kyllonen and Christal (1990) sought to understand how a general information processing factor was related sub-stratum variables from a hierarchical model of intelligence, Stauffer et al (1996) investigated how an overall ‘g’ factor predicts specific components of the information processing model. These authors suggested that individual differences in cognitive processes are driven by psychometric ‘g’.

Stauffer et al's (1996) results suggested that the measurement of human ability whether by traditional paper and pencil testing or by measuring cognitive components yields a measure of the same general mental ability construct. Given this finding and the findings discussed above (e.g., that aspects of the information processing system predict performance in a multitasking environment), it follows that any construct valid psychometric measure of cognitive ability should also predict performance in a multitasking environment.

Hypothesis 4 – A psychometric cognitive ability test (reading comprehension) is a strong predictor of performance in a multitasking environment.

Note. To interpret effect sizes in the social sciences, Cohen (1988) outlined the following guidelines: strong predictors have a magnitude of $r > .50$, moderate predictors have a magnitude between $r = .30$ and $r = .50$, and weak predictors have a magnitude between $r = .10$ and $r = .30$. The predicted magnitudes of the relationships proposed in the hypotheses in this study follow this convention.

<u>Polychronicity</u>

Polychronicity is another individual difference variable that may explain differences in performance in a multitasking environment. Polychronicity has not received a lot of attention as a predictor of job performance, but it has explained important variance in job performance in a few recent studies (e.g., Conte & Jacobs, 2003; Conte & Gintoft, 2005). That said, some other recent studies have failed to find a relationship between polychronicity and multitasking (Ishizaka et al, 2001; Konig, et al, 2005).

The concept of polychronicity was introduced by Hall (1959). Hall (1959) and in subsequent writing (e.g., Hall, 1988) distinguished between two types of organizing time.

Hall said that people organize events as distinct items or people organize events in such a way that they attempt to do multiple things at one time. The former approach is called monochronic and the latter is called polychronic (Bluedorn & Denhardt, 1988). Cotte and Ratneshwar (1999) added that monochronic people view time linearly; they take a "one at a time" approach. Alternatively, polychronic people use time for many purposes at once; in other words, they multitask. Conte, Rizzuto, and Steiner (1999) described monochronicity/polychronicity as a person's preferred temporal pattern of behavior. Cotte and Ratneshwar (1999) point out that perceptions of time are influenced by three factors: (1) culture, (2) social networks, and (3) personality individual differences. As will be discussed later, there does seem to be some confusion around whether polychronic preferences are an individual difference or a situational consequence. Treating it as an individual difference variable has received empirical support (e.g., Kaufman-Scarborough & Lindquist, 1999; Bluedorn, Kaliath, Strube, & Martin, 1999; Conte et al, 1999; Conte & Jacobs, 2003). Most authors suggest that monochonicity/polychronicity makes the most sense operationalized as a "trait" rather than a "type"; that is, monocrhonicity/polychronicity exists along a continuum rather than a dichotomous classification (Kaufman-Scarborough & Lindquist, 1999; Benabou, 1999; Conte & Jacobs, 2003).

Research on polychronicity uses the term in several different ways. Understanding the different usage of the term further illustrates what this construct is. Authors refer to "polychronicity", "polychronic behavior", and "polychronic time". Often these terms are used interchangeably, but they have subtle differences that are important to understand relative to their relation to multitasking environments.

"Polychronic behavior" seems to be used as a synonym for multitasking. For example, Bluedorn et al (1999) suggest that polychronic behavior can refer to doing multiple things simultaneously or the active interspersing of and dovetailing of several activities within the same general time period. Clearly, from this explanation, it follows that polychronic behavior equals multitasking. "Polychronic time" seems to refer to the environment; that is, polychronic time is a synonym for a multitasking environment. For example, Kaufman-Scarborough & Lindquist, 1999 suggest that polychronic time is characterized by overlaps of activities, interruptions, and the dovetailing of tasks. Benebou (1999) states that polychronic time is characterized by events occurring unpredictably and simultaneously. Finally, Bluedorn et al (1999) describe polychronicity as the preference for polycrhonic time and the belief that polychronic behavior is the best way to do things. In other words then, polychronicity is the preference for multitasking situations and the belief that multitasking strategies are the best way to complete tasks.

Several authors have commented on the usefulness as a predictor of job performance of this preference variable. Conte et al (1999) suggested that preferences for time use are important factors that can determine success in the workplace. These authors suggest that levels of polychronicity explain "how" people approach their work and they called for future research on how polychronicity predicts performance across a variety of work settings and industries. Kaufman-Scarborough & Lindquist (1999) suggest that polychronicity is present to varying degrees in the workplace, even among workers who are "side-by-side". Thus, meaningful individual differences can be identified and prediction can be researched.

While the relationship between polychronicity, polychronic time, and polychronic behavior seems obvious, there is some disagreement and confusion around what exactly polychronicity is, or at least at what level it is most appropriately operationalized. Some people operationalize polychronicity as a cultural characteristic (e.g., Bluedorn & Denhardt, 1988), others suggest it is a personality trait (e.g., Benabou, 1999), others suggest that it is not a personality variable, but rather a relatively stable preference (e.g., Persing, 1999). Still others refer to polychronicity as a task orientation or strategy rather than a preference for a situational characteristic (e.g., Moustafa, Bhagat, & Babakus, 2005). Cotte and Ratneshwar (1999) conducted a large-scale qualitative study about what drives polychronicity. They found support for their assertion that polycrhonic behavior depends on individual differences separate from culture and social networks. Several authors refer to polychronicity as an individual's temporal personality (e.g., Benabou, 1999) and Conte and Gintoft (2005) found that polychronicity is a relatively stable individual personality characteristic that is distinct from the five factor model.

Given that the most research up to this point suggests that there are important cultural and social differences in polychronicity (Hall, 1983) and given that the five factor model reliably generalizes across cultures, it is doubtful that polychronicity is a "true" and distinct personality trait. That said, researchers have found meaningful individual differences across and within cultures and work groups. While the exact nature of this preference for multitasking has yet to be understood, recent research has identified aspects of the nomological network surrounding polychronicity.

While research on polychronicity as an individual difference variable is still relatively new, there have been studies that have reported the relations between polychronicity, extraversion, agreeableness, intellectance, neuroticism, conscientiousness, Type A behavior pattern (TABP), cognitive ability, and job performance criteria. Conte and Jacobs (2003) investigated the relations between polychronicity, several common predictors of performance, and job performance in a train operator sample. They found polychronicity correlated with several components of TABP with a modest magnitude. Specifically, they found that polychronicity related to time urgency $r = .16$, achievement striving $r = .18$, and impatience/irritability $r = .18$. These authors suggest that the relatively weak correlations suggest that polychronicity is distinct from TABP. These authors also found that polychronicity was not strongly correlated with the big five. Specifically, these authors observed the following relationships with polychronicity: Extroversion $r = .21$, Conscientiousness $r = -.15$, Nueroticism $r = .05$, Intellectance $r = .07$, and Agreeableness $r = -.07$. Lastly, Conte and Jacobs observed a correlation between polychronicity and cognitive ability of $r = .15$. From these data, polychronicity appears to be a relatively distinct predictor. Conte and Jacobs (2003) reported that polychronicity did contribute incremental validity beyond these predictors in their sample of train operators.

Conte and Gintoft (2005) found similar results. These researchers found that polychronicity predicted customer service $r = .23$, sales performance $r = .22$, and overall performance $r = .23$ in a sample of computer sales employees. This study replicated the Conte and Jacobs' (2003) finding that polychronicity explains unique performance variance beyond the big five. In this study, polychronicity was related to extroversion $r = .22$, similar to the earlier Conte and Jacobs (2003) study. Unlike Conte and Jacobs (2003), conscientiousness was not related to polychronicity. Conte and Gintoft (2005) suggest that polychronicity and extraversion are related due to the tasks involved with dealing with developing networks and seeking out information. Hall (1983) said that polychronics are more likely than monochronics to be relationship oriented. While much more research is needed these studies suggest that polychronicity is relatively distinct from more traditional predictors used in selection settings, but also that it explains incremental validity in performance criteria.

While polychronicity has predicted job performance in a small handful of studies, there is relatively little empirical evidence suggesting that it predicts performance in demanding multitasking environments. There are however, some theoretical links to consider. Most of this line of thinking has to do with how polychronics are better suited to perform in situations with salient and demanding time pressures, which is one of the key components of the multitasking environment as discussed above. For example, Kaufman-Scarborough & Lindquist (1999) theorized that polychronics were more likely to be able to be comfortable with interruptions and actively switch tasks than are monochronics. These authors cite Gross (1987), who found that a polychronic style impacts perceptions of time pressures and the amount and order of time spent on tasks. Kaufman-Scarborough & Lindquist (1999) reasoned that being faced with changing circumstances brings stress to a situation if an individual is not comfortable with change. Since polychronics prefer this type of change they are less likely to perceive a multitasking environment as a stressor. In the Kaufman-Scarborough & Lindquist (1999) study, polychronics were more comfortable switching tasks and dividing attention. These authors as well as Cotte and Ratneshwar (1999) concluded that organizational outcomes are enhanced by maximizing the fit between the environment and the preferred level of multitasking. In other words, these authors suggest organizations can see performance improvements by hiring polychrons into demanding multitasking environments. While this hypothesis has not been supported, Conte and Jacobs (2003) did find support that monochrons outperform polychrons in a monochronic environment. Bluedorn et al (1999) suggest that the performance differences are driven by a polychron's ability to anticipate multitasking; they plan on having to engage in frequent task switches, so when confronted with them, they are ready.

While there is a growing body of theoretical information suggesting that polychronicity will predict performance in multitasking environments, there are also some theoretical arguments suggesting that perhaps it may not. In the Kaufman-Scarborough & Lindquist (1999) work referenced earlier, there were points that suggest that polychrons may not have an advantage in multitasking environments. These authors found that monochrons and polychrons did not differ in their ability to prioritize activities. They also found that they did not differ in understanding which task to switch to next. Persing (1999) mentioned the role of personal agency in the relationship between polychronicity and performance in a multitasking environment. Persing (1999) suggests that the power of the situation can trump individual differences in polychronicity. This authors suggests that polychrons can behave like monochrons and vice versa if the situation calls for it. This author believes that polychronicity is dependent on personal agency, which means that the preference for multitasking will only exist if the individual mandates how the multitasking will occur in the given environment. The authors posit that even polychrons will not exhibit a preference for polychronic time if the multitasking demand is driven situationally or externally.

This line of thinking is critical to understanding whether or not polychronicity will predict performance in a multitasking environment in two ways. First, Persing's (1999) ideas suggest that if the situation defines the multitasking demand, people can rise/fall to the occasion despite their position on the polychronicity continuum. While monochrons may not prefer to multitask, Persing (1999) suggests that this does not mean that they cannot multitask. Second, the author's idea that polychronicity depends how the multitasking is created in the situation also questions whether or not polychronicity will predict performance in a multitasking situation. If Persing (1999) is correct, a preference for multitasking will not emerge, even among typically polychronic people. If the situation defines the necessary degree of multitasking, then polychronicity may not be an important predictor of performance in multitasking envrironments. In most entry-level demanding multitasking situations, such as the situation created in the job-relevant multitasking situation used in this study, people do not have much control over the degree of multitasking that is needed. Thus, Persing's ideas would predict that polychronicity would not be an important predictor of performance in these types of multitasking environments.

In support of Persing's ideas, two recent studies found that polychronicity is not an important predictor of performance in multitasking environments similar to the one under study in this project. Ishizaka, Marshall, and Conte (2001) created a multitasking simulation where participants had to engage in frequent tasks switches among several simultaneously presented tasks that involved completing math problems and manipulating "gauges" on a computer screen. There was also an auditory tasks that required participants to recall sets of words that were presented during the simulation.

The participants in this study consisted of 118 undergraduate psychology students. The magnitudes in the relationships between polychronicity and the criterion measures were typically in the low- to mid-teens. With low power from the relatively small n, these weak relationships were not statistically significant, thus the authors concluded that polychronicity was not predicting performance in the multitasking simulation.

In another recent study, Konig et al (2005) used the SIMKAP, which is a research-oriented multitasking simulation to address the relationship between several predictors, including polychronicity and performance in a multitasking environment. These researchers had 131 college students participate in the study. Like Ishizaka et al (2001), these authors did not find a statistically significant relationship between polychronicity and the multitasking criteria. The magnitude of the relationships, while not statistically significant were similar to Ishizaka et al. (2001); they averaged $r = .11$ across the three SIMKAP scores.

While the authors of these studies appropriately concluded that they did not observe a meaningful influence on performance from polychronicity, there may be some prediction happening. Both studies observed relationships trending in the predicted direction, but they did not reach statistical significance. Both groups of researchers predicted that polychronicity would predict multitasking based on the literature addressed above. Perhaps the relative lack of relationship was due to the ideas of personal agency and situational strength discussed by Persing (1999).

In the current study, there is considerably more statistical power to detect an effect (n is greater than 2,000). Also, the participants in this study consist of actual job candidates that may be more motivated than college students participating for course credit. The increased “stakes” may heighten the stressors of the multitasking situation that have been predicted to drive the relation between polychronicity and multitasking performance. Thus, the relation between polychronicity and performance in a multitasking environment may be more interpretable in this research. That said, given the Persing’s (1999) ideas and the finding of the Ishizaka et al. (2001) and Konig (2005) studies, this relationship is expected to be weak.

Hypothesis 5 – Polychronicity is a weak predictor of performance in a multitasking environment.

Type A Behavior Pattern

Another possible non-cognitive predictor of performance in multitasking environment is Type A Behavior Pattern (TABP). TABP was described initially by Friedman and Rosenman (1959). These authors noticed coronary heart disease patients tended to be highly competitive, achievement striving, aggressive, and impatient. In this initial work and subsequent writings, these researchers defined TABP as an “action-emotion” complex. They explained that Type A’s are “aggressively involved in a chronic, incessant struggle to achieve more in less and less time and if required to do so against the opposing efforts of other things and people” (Friedman & Rosenman, 1974; pp 67).

Since this pioneering work by Friedman and Rosenman, lots of research has been conducted to explore the concept of TABP. Much of the initial research on TABP focused on its links to physical symptoms, such as heart disease. Matthews and Brunson (1979) commented that TABP had become widely known for its effects on cardiovascular health. They reported that Type A's have double the risk of heart disease relative to Type B's (Type B is typically defined as the absence of Type A characteristics; although, Type A and Type B are conceptualized as opposing poles on a continuum (Moch, 1984)). Further, Bermudez, Perez-Garcia, Sanchez, and Elvira (1990) reported that this behavioral style is a risk factor for heart disease on par with high cholesterol, high blood pressure, and smoking.

Over the years, lots of studies have outlined the identifying characteristics of TABP. Common themes among the characteristics in these studies are present. For example, studies on the topic have typically characterized those with Type A as ambitious, competitive, extreme sense of time urgency, impatient, irritable, aggressive, hostile, use rapid explosive speech, focus on quantity over quality, etc. (e.g., Burnham, Pennebaker, & Glass, 1975; Carver & Glass, 1978; Matthews & Brunson, 1979; Moch, 1984; Spence, Helmreich, & Pred, 1987; Bermudez et al., 1990; Landy, Rastegary, Thayer, & Colvin, 1991; Schaubroeck & Williams, 1993; DelaCasa, Gordillo, Mijias, Rangel, & Romero, 1998; Ishizaka et al., 2001).

As this varied collection of characteristics presumes, TABP is not a single construct, but rather a combination of component constructs. Even though the component parts of TABP appear to be clearly distinct, much of the TABP research has used global measures of TABP. In the last 20 years, several researchers have pointed out that this global approach to TABP has methodological and conceptual weaknesses (Spence, Helmreich, & Pred, 1987; Helmreich, Spence, & Pred, 1988; Bluen, Barling, & Burns, 1990; Edwards et al., 1990; Ishizaka et al., 2001).

In a series of studies (Spence, Helmreich, & Pred, 1987; Helmreich, Spence, & Pred, 1988), researchers Spence, Helmreich, and Pred found that TABP could be split into two distinct constructs: Achievement Striving and Impatience/Irritability. These authors suggested that the achievement striving aspects of TABP drives positive peformance outcomes of TABP; while the impatience/irritability aspects drive the negative health outcomes of TABP. Bluen et al. (1990) explained that the achievement striving portions of TABP reflect the extent to which people take their work seriously, are active, and work hard. Alternatively, Impatience/Irritability reflects intolerance, anger, and hostility. In the studies by Spence, Helmreich, and Pred (1987; 1988) and Bluen et al. (1990), the researchers were able to find differential relationships with these two factors of TABP and various outcomes: Achievement Striving predicted academic performance, publications, sales performance, and job satisfaction; Impatience/Irritability predicted depression, general health, and complaints about physical symptoms. The results of these studies suggest that 1) using global TABP measures obscures differential relationships with different criteria; and 2) the achievement striving aspects of TABP are likely the most important to investigate in research on TABP and job performance.

Spence et al. (1987) pointed out that generally speaking Type A's outperform Type B's (but have more health issues). These performance differences could be observed due to a number of different performance characteristics displayed by Type A's. Carver, Coleman, and Glass (1976) found that Type A's will work at a maximal rate even when there is no time deadline. Further, Moch (1984) and Delbridge (2000) pointed out that Type A's cope with stress differently. Moch (1984) found that Type A's focus on the experience of stress less while performing a task, even though they show other signs of arousal (e.g., increased pulse rate and blood pressure). Landy et al (1991) suggested that Type A's are more acutely aware of time and tend to schedule events in shorter periods of time. Matthews and Brunson (1979) found that when Type A's anticipate a loss of control over their environment, they make greater initial efforts. In other words, they attempt to accomplish as much as possible while they still have the chance. Finally, Bermudez et al. (1990) suggests that Type A's will allocate more attention to tasks when they perceive the task to be personally challenging.

These performance characteristics suggest that Type A's will likely outperform Type B's in a variety of settings, particularly in multitasking environments. Given that Type A's are prone to work fast, focus on their control of the environment, focus on stress less than Type B's, and have an acute awareness of time pressures, Type A's seem tailor-made for multitasking environments.

While these characteristics do suggest a clear relationship between TABP and performance in a multitasking environment, research on differences in resource allocation between Type A's and Type B's suggests that this relationship may not be straightforward. Matthews and Brunson (1979) suggested that Type A's and Type B's differ in how they allocate attention. They explained that due to the chronic desire to achieve, Type A's focus all of their attention to the tasks at hand. The authors predicted that Type A's would not notice peripheral events that are not relevant to the performance of the immediate task. Rather, these researchers predicted that Type A's would only focus on the environmental stimuli that would aid them in performing the present task.

To test this hypothesis, Matthews and Brunson (1979) used a dual task paradigm. They expected that Type A's would allocate less attentional resource to the secondary task than Type B's. The participants were given a global TABP measure and then introduced to the dual tasks. They were allowed to practice the primary task before the experimental trials began. The secondary task was mentioned as an "after thought". The researchers found that Type A's reaction times were much slower for the secondary task than Type B's, however Type A's performed more accurately on the primary task. This finding suggests that because of the intense concentration on the current task, Type A's did not perform well in the dual task multitasking environment.

Moch (1984) drew similar conclusions. In this study, the researcher hypothesized that noise interruptions would be more problematic for Type B's than Type A's when working on complex tasks. To test this hypothesis, 104 participants were exposed to high frequency tones while performing two working memory tasks. The level at which the participants noticed the noise disturbance was recorded as their noise sensitivity. They found that Type A's maintain their noise tolerance regardless of the task, while Type B's are more sensitive to these noise interruptions. While this difference suggests that Type A's are able to drown out interruptions and keep more attention on the task at hand, it could also indicate that Type A's are less sensitive to the cues to switch tasks in multitasking environments.

Bermudez et al. (1990) pointed out that Type A's feel less fatigue and other physical symptoms while working on involving and challenging tasks, but these differences disappear upon task completion. These authors suggested that this finding is due to resource allocation differences. They point out that Type A's only focus on elements relative to the task at hand, while ignoring other distractions from the environment. They tested these ideas with 18 Type A's and 18 Type B's. The participants were presented with shapes in a tachistiscope. Participants had to determine whether or not shapes matched what they were instructed to look for. In the divided attention condition, the participants had to make multiple determinations for multiple shapes presented simultaneously. The results of this study contradicted the work of Matthews and Brunson (1979) and Moch (1984). These authors found that when participants were instructed to divide attention, Type A's could do so faster and more accurately than Type B's.

Building on this research, DelaCasa et al. (1998) suggested that Type A's use different cognitive strategies relating to attentional and memory processes. One such difference is that Type A's adopt strategies that make them better able to focus on task-relevant aspects of environments, as was found by Matthews and Brunson (1979) and Moch (1984). DelaCasa et al. (1998) pointed out that the situation in these studies was one where the primary task was unambiguously the most important task and, thus, this strategy was appropriate.

In the Bermudez et al. (1990) study, a strategy focusing on a single task was not appropriate. It was made clear to participants that attention needed to be divided across tasks, thus, the Type A's adopted a different cognitive strategy than they did in the earlier studies. DelaCasa (1998) suggested that, rather than focusing on a single task, Type A's make sustained efforts to achieve the best results in performance of all tasks when situations are ambiguous or if they feel that all tasks "matter". These authors suggested that in ambiguous and complex multitasking environments, Type A's engage in "hypervigillance". This approach allows them to process the greatest amount of information possible.

Ishizaka et al. (2001) replicated these findings in a simulation designed specifically to measure performance in a multitasking environment. These authors found that the achievement striving and competitiveness aspects of TABP predicted performance in a complex three-task multitasking situation and in ambiguous situations. These authors suggest that employees high in achievement striving may be successful in work environments where prioritization of multiple tasks is appropriate.

The finding from this series of studies suggest that Type A's outperform Type B's in multitasking environments in three ways. First, they perform primary tasks better in dual task situations, due to more focused attention. Second, they perform tasks better in the face of interruptions because of their ability to suppress task-irrelevant external stimuli. Lastly, they perform better in complex or ambiguous multitasking simulation because of their hypervigilant approach. These findings in concert with findings suggesting that Type A's have a more acute sense of time urgency (Landy et al, 1991), suggest that the achievement striving and competitiveness components of TABP will be moderate predictors of performance in job-relevant multitasking environments.

Hypothesis 6 – Achievement Striving is a moderate predictor of performance in a multitasking environment.

Hypothesis 7 – Competitiveness is a moderate predictor of performance in a multitasking environment.

Big Five Personality Traits

While polychronicity has been conceptualized as a personality trait by some researchers and TABP consists of a collection of traits, to really understand whether or not personality predicts performance in a job-relevant multitasking environment the Five

Factor Model of personality should be tested. In the current research, each of the "big five" personality traits will be tested to assess whether or not they are predictors of performance in complex multitasking situations. While it is not expected that all of the five factors will predict performance in the simulation, it will be useful to collect evidence about the relative position of each trait in the nomological network. Given that there have been very few tests of the prediction of the big 5 specifically in multitasking environments, even the identification of non-relations is quite informative.

The idea that personality can be explained using five factors has been around for quite some time. In 1949, Fiske was perhaps to be the first to find five distinct factors, in the researchers attempt to replicate Cattell's 16 factors. Fiske's five factors were labeled social adaptability (similar to extraversion), conformity (similar to agreeableness), emotional control (similar to neuroticism), inquiring intellect (similar to intellectance or openness to experience), and will to achieve (similar to conscientiousness; this factor from Fiske's data came to light in a reanalysis by Digman (1990). At the time, Fiske's (1949) work did not catch on and research on these factors was slow to follow.

Later, Tupes and Christal (1961) also found five factors. Tupes and Christal's five factors were surgency (similar to extraversion), agreeableness, dependability (similar to conscientiousness), emotionality (similar to neuroticism), and culture (similar to intellectance). These findings were published in an obscure Air Force technical report and thus did not receive much attention. In 1963, Norman came across the technical report and replicated the five factors. Norman (1963) is credited with being the first to use the commonly accepted labels of the big five. In fact, the big five has been referred to as Norman's Big Five. Barrick and Mount (1991) reported that Norman's Big Five (extraversion, emotional stability, agreeableness, conscientiousness, and culture) hold up across theoretical frameworks (e.g., Goldberg, 1981), different instruments (Costa & McCrae, 1988), and different cultures (Bond, Nakazato, & Shiraishi, 1975; Noller, Law, & Comrey, 1987; Saucier & Ostendorf, 1999).

Given the robust nature of these five factors of personality in normal human populations and given that in study after study these five traits differentiate between people, manifest themselves lawfully across situations, and endure over time, the five factor model is an important consideration in any study of individual differences. If the goal is to understand individual differences predictors of performance in any situation, the big five should always be investigated. In the sections that follow, each of the big five traits will be discussed in relation to performance in a multitasking environment and hypotheses will be suggested about the likely relationships.

<u>Extraversion</u>. According to Costa and McRae (1988) extroverted people are warm, outgoing, gregarious, assertive, energized, and excitement seeking. Extraversion is not simply talking a lot or being outgoing. Rather extroverts are risk-takers who seek stimulation. Extroverts tend to have a lot of friends; whereas introverts have fewer, but closer friends. Barrick and Mount (1991) found that extroversion predicted job performance in jobs that required a lot of social activity.

While not much research has been done investigating the relationship between personality and performance in multitasking environments, extraversion has probably received the most attention. There are several theoretical connections between extroversion and performance in a multitasking situation. Dickman (1985) suggested that extraverts and introverts have different information processing characteristics; extraverts focus on speed, whereas introverts focus on accuracy. In a multitasking environment you may expect introverts to do less but do it accurately, while extraverts may do more but their work could potentially be of lower quality. In a test of this hypothesis, Dickman and Meyer (1988) tested people high in impulsivity, which is similar to extraversion (see Gray, 1987), to see if they emphasize speed at the expense of accuracy. They found that high impulsives were faster and less accurate.

Lieberman and Rosenthal (2001) suggested that the difference in social interactions between introverts and extraverts is based on the extroverts' ability to perform interpersonal multitasking. That is, in the course of a social interaction, meaning and messages are communicated through many verbal and non-verbal cues. Extraverts are better able to attend to and interpret the breadth of simultaneously occurring cues. They found support for this idea in a series of studies. These authors suggested that the root of this relation has to do with the levels of chatecholamines in the prefrontal cortex. Lieberman and Rosenthal stated that chatecholamines are important for multitasking as well and that good multitaskers should have similar levels to extraverts, thus suggesting a positive relationship between multitasking performance and extraversion.

Further, arguments for the influence of extraversion on multitasking can be based in activation theory (Gardner, 1986; Gardner & Cummings, 1988). These researchers stated that if the arousal level in a situation is too low, then people will become bored and performance will suffer. Conversely, if the arousal level is too high, then the situation will be anxiety provoking causing detriments in performance. Consistent with activation theory, Matthews, Jones, and Chamberlain (1989) and Matthews, Davies, and Lee (1990) explained extraversion is associated with individual differences in attention. For example, introverts are more susceptible to adverse effects of increased processing demands; whereas, extraverts will perform best under high levels of stimulation. In other words, relative to their introverted counterparts, extroverts get bored more easily and can tolerate higher levels of complexity in any given situation. From this activation theory argument, it follows that an extravert's ideal situation is a complex multitasking environment; simple mono-task environments would not be arousing enough to keep them interested. Conversely, introverts are more likely to become anxious and experience stress in multitasking environments and would prefer situations with fewer stimuli.

From these theoretical arguments it seems clear that extraversion is correlated with performance in a multitasking environment, however, this relation eluded researchers in the one known test of this hypothesis. Konig et al. (2005), found non-significant weak correlation coefficients between extraversion and multitasking performance across three multitasking criteria ($r = .13$, $r = .11$, and $r = .09$). The sample size in this study was small, which can explain the lack of significance, but even so the strength of this relationship is surprising given the theoretical rationale for the hypothesis. Konig et al. (2005) acknowledge their surprising counter-theoretical results and suggest that replication is needed. Towards this end, the hypothesis that extraversion is a predictor of multitasking performance will be tested.

Hypothesis 8 – Extraversion is a weak to moderate predictor of performance in a multitasking environment.

Neuroticism. Neuroticism or emotional stability or adjustment is characterized as the tendency to have mood swings or repeated bouts of negative affect. Costa and McCrae (1988) explained that people high in neuroticism are often anxious, depressed, self-conscious and easily hurt. For the sake of this study, neuroticism is defined as tolerance to stress (the host organization uses a narrow measurement of this trait, due to the belief that only the stress tolerance sub-scale is job-relevant). Therefore, despite some of the theoretical rationale for the predictive validity of neuroticism as a whole, this study focuses only on stress tolerance.

While little has been done investigating the link between stress tolerance and multitasking performance, there is reason to think such a relationship may exist. Delbridge (2000) points out that research on stress suggests that stress tolerant people may be better equipped to deal with the feeling of urgency created by the demanding time pressures in a multitasking environment. Carver, Peterson, Follansbee, and Scheier (1983) found that highly anxious people exhibit lower persistence on difficult tasks. These lines of thinking suggest that if highly stress tolerant people are less prone to experience stress in time urgent situations and that if the experience of stress is related to task persistence and withdrawal, then highly stress tolerant people are likely to perform better in a multitasking environment.

While this relationship makes intuitive sense, this effect is probably moderated by other variables that are not measured in the current study. For this hypothesis to hold true, the participants must interpret the multitasking situation as stressful to some degree or another. Whether or not this perception exists could depend on variables such as tolerance for ambiguity, locus of control, feelings of self-efficacy, etc. Because these constructs are not measured by the host organization, it is impossible to know if they are impacting the relationship between stress tolerance and performance in a multitasking environment. Given the likelihood of the presence of such moderators, it could prove difficult to uncover the main effects of stress tolerance on performance in a demanding multitasking environment, therefore, only a weak relationship is predicted.

Hypothesis 9 – Stress tolerance is a weak predictor of performance in a multitasking environment.

Intellectance. Of the big five, perhaps the most disagreement has surrounded this factor. Intellectance is also often referred to openness to experience. Also, the term culture has been used to label this trait. According to Costa and McCrae (1988), people high in intellectance are more curious and their lives are experientially richer. They are more willing to entertain novel ideas and unconventional values. People low in intellectance are more conventional in behavior and conservative in outlook. Intellectance is associated with education and intelligence but is a distinct construct.

While no empirical studies have addressed the issue of whether or not intellectance predicts performance in multitasking environments, there are some reasons to think that it probably does. Barrick and Mount (1991) and Mount and Barrick (1995) point out that people high in intellectance tend to be imaginative, curious, and more open to new experiences and change. McCrae (1996) suggested that people high in intellectance tend to have more behavioral flexibility. These two ideas are relevant to performance in a multitasking environment. Because performance in a multitasking environment requires frequent task switches, people that are more well suited to change may have an advantage over people that are resistant to change. Further, because people high in intellectance value newness and change and have the behavioral flexibility to deal with it, they may perceive less uncertainty or, as Delbridge (2000) suggested, they may be less bothered by the uncertainty inherent in a multitasking environment. Therefore, it is hypothesized that because of the high need for change and the behavioral flexibility that characterizes people high in intellectance, these people will perform better in a multitasking situation than those low in intellectance.

Hypothesis 10 – Intellectance is a moderate predictor of performance in a multitasking environment.

Conscientiousness. Conscientiousness refers to an individual's degree of organization, persistence, and motivation in goal-directed behavior (Costa & McCrae, 1988; Barrick Mount, 1991). On the high end, conscientious people are fastidious, dependable, careful, planful, and thorough. On the extreme low end, they are lackadaisical and sloppy.

Barrick and Mount (1991) found that conscientiousness does predict job performance in a large-scale meta-analytic study. That said, these authors have suggested in later writings that this relationship may depend on the job. For example, these authors suggested that the relationship between conscientiousness and job performance is likely more prevalent for jobs that allow a high degree of autonomy. In the type of multitask situation presented in this study, there is very little autonomy; the tasks to work on and how to work them are defined by the situation (this is also the case for the entry-level jobs that the simulation mimics). In the end, being organized, dutiful, and disciplined is likely to do little to impact a person's task performance in a multitasking environment. It is possible that conscientiousness may still be important for jobs with a high demand for multitasking for reasons other than task performance. For example, people high in conscientiousness may be more dependable and exhibit citizenship behaviors. However, it is not likely that the level of conscientiousness will impact their performance when it comes to accomplishing multiple task goals in the same general time period. In the current study, no relationship is expected between conscientiousness and performance in a multitasking environment.

Hypothesis 11 – Conscientiousness does not predict performance in a multitasking environment.

Agreeableness. Agreeableness refers to the quality of one's interpersonal orientation. It can be thought of as existing along a continuum with antagonism on the low end and compassion and caring on the high end. Someone high in agreeableness is altruistic, helpful, and trusting; whereas, someone low in agreeableness is egocentric, skeptical of others, and disingenuous (Costa & McCrae, 1988; Digman, 1990). Agreeable people tend to be helpful and supportive. People low in this trait tend to be aggressive and prone to conflict. Agreeableness is more relevant to social situations than task situations. While agreeableness is likely to impact how people are perceived, and perhaps rated, in their jobs, it is not likely to impact actual task performance. To this date no theories have emerged that suggest that someone's level of agreeableness will impact how quickly and accurately, they are able to switch between non-social tasks. Thus, agreeableness is predicted to have no relation to performance in a multitasking environment.

Hypothesis 12 – Agreeableness does not predict performance in a multitasking environment.

The individual difference hypotheses are listed in Table 3.

Table 3

Proposed Individual Difference Hypotheses

#	Hypothesis
4	A psychometric cognitive ability test (**reading comprehension**) is a **strong predictor** of performance in a multitasking environment.
5	**Polychronicity** is a **weak predictor** of performance in a multitasking environment.
6	**Achievement Striving** is a **moderate predictor** of performance in a multitasking environment.
7	**Competitiveness** is a **moderate predictor** of performance in a multitasking environment.
8	**Extraversion** is a **weak to moderate predictor** of performance in a multitasking environment.
9	**Stress tolerance** is a **weak predictor** of performance in a multitasking environment.
10	**Intellectance** is a **moderate predictor** of performance in a multitasking environment.
11	**Conscientiousness does not predict** performance in a multitasking environment.
12	**Agreeableness does not predict** performance in a multitasking environment.

METHODS

Participants

The participants in this study consist of applicants applying for entry level call center positions with one of the nation's leading automotive insurance providers. Each participant was applying for a job in one of the ten regional call centers operated by the host organization. These call centers are in Buffalo, NY; Woodbury, NY, Fredericksburg, VA; Virginia Beach, VA; Macon, GA; Lakeland, FL; Coralville, IA; Dallas, TX; Tucson, AZ; and San Diego, CA. The data set for this study is comprised of 2,711 job candidates. The sample is 63% female and 37% male. The age range in this sample is 18-years-old to 83-years-old. The mean age of the participants is 29-years-old with a standard deviation of 9 years.

All of the data was collected as part of the host organization's typical selection process. This step in the process occurs at the beginning of the selection system. To get to this stage, the job candidate must pass an application review and a 10-15 minute pre-screening interview covering basic job requirements (e.g., legal age, required education, job history, etc.). These two steps have a pass rate of about 70%, but, even with these liberal yield rates, the cuts to the candidate pool during these steps does create some restriction of range relative to the general population. That said, this range restriction also makes the sample that much more comparable to the population of people that are currently holding positions similar to the simulated position. In other words, the range restriction makes this simulation data more generalizable to the target jobs.

Measures

The multitasking simulation discussed in detail in the Appendix serves as the dependent measure in this study. The purpose of this tool is to place candidates in a multitasking environment similar to the multitasking situation on the actual job and ask the candidates to complete similar tasks as those that are completed on the job. The tool is used to discriminate between candidates' performance in this job-relevant multitasking environment.

As illustrated in Table 1, the simulation includes two scored trials consisting of eight sets of tasks in each. Each set of tasks is called a query. There are phone queries and email queries. Both types of query requires the candidate to extract ten points of information and input it across a series of computer tabs and screens using mostly drop down menus and a few text fields. In the performance of each query 14 data points are collected. These include whether or not the query was opened, the time the query was opened, ten accuracy measures based on how the candidate works the query, whether or not the query was completed, and the time the query was completed. Currently, the host organization only uses a simple aggregation of the accuracy measures when making decisions about job candidates, but all of the data is available for this dissertation. At the end of the simulation, the tool collects 224 pieces of performance data about each candidate.

As discussed previously, several individual difference constructs are hypothesized to predict performance in a multitasking environment. The tools used to measure these constructs are explained below.

To measure cognitive ability, the host organization's cognitive ability test will be used. The test is called the Basic Skills Test (BST) and is a measure of reading comprehension. The test presents the candidate with reading passages followed by sets of questions about the content of the reading passage. There are 36 items on the test about four reading passages. The test takes approximately 20 minutes to complete. The host organization has been using this tool for 7 years. In several validation studies, this test has been shown to make valid inferences about success in training and on state mandated property and casualty insurance licensing exams. The host organization feels that this tool is their best predictor of early-tenure success on the job.

The patterns of scores observed from the BST have shown evidence of construct validity in past research. Training test performance and licensure exam performance are two of the key criteria the BST is used to predict. Both of these exam types require associates to learn, recall, and apply many complex insurance-related principles and regulations, as well as, organization-specific processes and tasks. These training and licensing exams have a heavy cognitive demand and the BST has shown to have strong moderate relations with these criteria (e.g., BST & licensing exams across two studies, $r = .42$; $r = .45$; BST & training final exams across three studies, $r = .33$; $r = .36$; $r = .44$). Further, the magnitude of relationships between the BST and the other predictors in this study have shown to be as would be expected in past research. Table 4 shows the correlation coefficients between the predictor measures proposed in this study from some recent research. The BST had a weak relationship with intellectance and polychronicity ($r = .18$ for both). All other predictor correlation coefficients are less than $r = .15$. The convergent validity evidence from the cognitive-loaded training and licensing tests and the discriminant validity from the personality measures suggest that the BST is construct valid.

Table 4

Convergent and Discriminant Validity of the Predictor Measures (From a Previously Conducted Study; N=1456)

Predictor Measures	1	2	3	4	5	6	7	8
1 Basic Skills Test	1.00							
2 Extraversion	.13	1.00						
3 Intellectance	.18	.56	1.00					
4 Conscientiousness	-.12	.39	.53	1.00				
5 Agreeableness	.06	.52	.58	.55	1.00			
6 Stress Tolerance	.07	.63	.59	.53	.56	1.00		
7 Competitiveness	.02	.31	.11	.07	-.04	.15	1.00	
8 Achievement Striving	-.09	.47	.49	.76	.40	.43	.22	1.00
9 Polychronicity	.18	.37	.42	.35	.41	.49	.04	.25

Note. All coefficients > $r = .05$ are significant at $p < .05$.

The remaining predictor measures come from the host organization's Personal Preferences Inventory (PPI). The PPI was developed in 2003, and consists of 118 items about behavioral tendencies. It is an in-house developed big 5 measure of personality with a few extra sub-scales. This tool has been administered to over 30,000 job candidates since being implemented in 2004. The reliability of the traits and sub-scales is well documented. The internal consistency reliability of these trait measures and sub-scales from two recent samples of PPI test takers is reported in Table 5. While most of the scales are at an acceptable level of reliability, the sub-scales that dip below conventional standard are comprised of fewer items, which impacts the sub-scales' observed reliability coefficients. Further evidence of construct validity can be taken from the predictor correlation coefficients reported in Table 4. The pattern of results across the sub-scales are as would be expected; some magnitudes are large likely due to common-method bias, but each of the five traits appear to be tapping independent variance.

Table 5

Internal Consistency Reliability for the Non-Cognitive Predictor Measures from Two Previously Conducted Studies

Scale	# of Items	Reliability results from an earlier Study (n = 2456)	Reliability results from another earlier study (n =1188)
Extraversion	**35**	**.91**	**.91**
Needs Attention	4	.82	.84
Assertiveness	9	.80	.82
Social Comfort	9	.81	.82
Fun Loving	6	.73	.73
Energy	7	.71	.73
Intellectance	**16**	**.81**	**.78**
Curiosity	5	.63	.64
Learning Orientation	7	.76	.74
Conscientiousness	**19**	**.85**	**.83**
Rule Abiding	5	.76	.75
Self Discipline	7	.69	.72
Agreeableness	**16**	**.85**	**.85**
Warmth	8	.80	.80
Sensitivity	8	.67	.69
Neuroticism/StressTolerance	**7**	**.75**	**.76**
Polychronicity	**4**	**.61**	**.64**
Type A Behavior (composite)	**13**	**.74**	**.75**
Achievement Striving	7	.68	.66
Competitiveness	6	.76	.78

Note. Trait measures are bolded and placed immediately above the sub-scales that comprise them.

The Type A scale and the Polychronicity scales used in this study had never been considered by the host organization prior to this study. These measures are comprised of items that were intended to measure different constructs, however, the content of these items suggest that considering them as Type A and Polychronicity, respectively, is appropriate. Achievement striving and competitiveness are included in the PPI as “extra” scales. As these scales are often considered important components of Type A measures of personality (e.g., Ishizaka, et al., 2001), they will be considered as such in this study.

The polychronicity sub-scale has received almost no attention from the host organization until this project. The items were not originally intended as a measure of polychronicity, but the content is similar to other polychronicity scales that have been shown to be construct valid (Conte, Rizzuto, & Steiner, 1999). To establish construct validity of the PPI polychronicity sub-scale, a pilot study was conducted in August 2006 using a supervisory version of the PPI, which included Bluedorn et al.'s (1999) ten-item measure of polychronicity as well as the four-item scale used in this study. Results from this supervisor-PPI study suggest that the polychronicity measure used in this study measures largely the same construct as the Bluedorn et al. scale. The relation between the two scales was $r = .58$ in a sample of 92 supervisors.

Procedure

The predictor and outcome data were collected as part of the hiring process of the host organization. The typical process is a multiple hurdle system involving the following selection components: recruiting, application review, screen interview, testing, "line" interview, role play, job observation, manager interview, and contingent job offer. The process typically takes 10 to 15 days (including weekend days) to move a successful candidate from application to job offer. The data being considered in this study comes primarily from the testing step in the selection system.

During the testing phase, the candidates take the BST, PPI, the Multitasking Simulation, and an experimental Grammar Skills Test (GST) that is currently not factored into the decision process. The order in which the tests are administered depends on the scheduling practices of each participating region. In general, about ¼ of the candidates will have taken each respective test first. Specific testing order is not identifiable for any given participant in the data set. All candidates included in the data set have scores on the Multitasking Simulation, BST, and PPI. The GST, interview, and role play data are currently not being considered in the study (but could be if it were deemed interesting).

Operationalizing Criterion Performance

Thus far performance on the simulation has been referenced in general terms. A lot of data is collected during the multitasking simulation, likewise, performance can be operationalized in several different ways. In general, these different potential performance variables are of three types: Accuracy criteria, completion criteria, and speed criteria. Further, each criterion measure can be calculated across three levels: the query level, the trial level, and the overall simulation level. Query level data and trial level data are most useful in testing the task characteristic hypotheses; whereas, overall simulation level data are most useful in testing the individual difference hypotheses. In the following paragraphs, each type of criterion variable is considered in terms of its definition and usefulness in shedding light on the hypotheses proposed in this study.

Accuracy Criteria. Accuracy criteria can be defined in two ways in this study: "total accuracy" and "accuracy attempted". Total accuracy refers to the number of items responded to correctly out of the total possible number of items. For example, there are nine accuracy items per query, 72 per trial, and 144 across the two trials. Total accuracy refers to the number of items answered correctly out of the complete set of tasks. This variable is perhaps the most important criterion. From an organizational perspective, this variable would be the key job performance metric. This variable is the best answer to the questions of whether or not the participants did what was needed despite the demands of the situation and whether or not they accomplished these tasks effectively. The merit of the hypotheses must be measured against this standard. In other words, if the individual differences predictors do not predict this criterion variable, it would be hard to say that they predict performance in a job-related multitasking environment. The descriptive statistics for this variable and the criterion intercorrelation coefficients are presented in Table 5. The descriptive statistics show plenty of variance and appear to be normally distributed.

Another accuracy criterion that appears to make sense on the surface is "accuracy attempted". In other words, it may make sense to understand performance by looking at how accurately participants answer the questions that they attempted. This variable is intuitively appealing because the nature of the multitasking simulation is designed to present the participant with more tasks than a participant can handle, so looking at how effectively the attempted items were handled could provide insight into performance. However, in generating the tasks, I was careful to make each individual item simple enough that when attempted most people would be able to perform accurately. The rationale behind this decision was to ensure that performance variance in the simulation was driven by a participant's ability to perform under the demands of a multitasking simulation rather than his or her cognitive ability to "figure out" the right answer. Because of this, the percentage of items correct out of the items that were attempted is quite high. This variable is of little use because it has very little variance (thus ensuring that the simulation is more than just a traditional reading comprehension or reasoning measure of cognitive ability). The mean for this measure was 93% with a standard deviation of only 8% (see Table 6) and due to this lack of variance and skew, this potential criterion variable shows the smallest relationships with the other criterion variables. Whereas, the total accuracy criterion is likely the most important criterion variable, the "accuracy attempted" criterion is not worth consideration in terms of hypothesis testing.

Table 6

Descriptive Statistics and Inter-correlation Coefficients for the Criterion Variables

Criterion Variables	N	Mean	SD	1	2	3	4	5	6
1 Total Accuracy	2711	.47	.26	1.00					
2 Accuracy Attempted	2655	.93	.07	.49	1.00				
3 Queries Opened	2711	13	3	.54	.44	1.00			
4 Queries Submitted	2711	9	5	.58	.39	.74	1.00		
5 Tasks Attempted	2711	71	38	.99	.40	.54	.57	1.00	
6 Avg. Reaction Time*	1071	19S	6S	.39	-.03	.41	-.43	-.31	1.00
7 Avg. Completion Time*	1071	45S	8S	.60	-.02	.63	-.45	-.53	.71

Notes. All coefficients > r = .05 are significant at p < .001. * = The correlation coefficients for reaction time and completion time were taken from a subset of the sample of people that completed more at least two thirds of the 16 queries. This approach was chosen to ensure stability in the average speed variables.

Completion Criteria. As alluded to above in consideration of the total accuracy variable, accuracy by itself is not enough; consideration of the tasks that are completed is also an indication of performance in a multitasking environment. There are three types of completion criteria that make sense to consider: queries selected, queries completed, and tasks completed. Each of these criterion variables indicate something about the number of tasks the participant either chose to complete or was able to complete. The number of queries opened indicates how quickly the participant was able to react to queries appearing in the queue and how quickly s/he was able to complete the previous query and move on to the next query. The number of queries completed indicates how many queries were submitted by the participant within the 60-second time limit that the participant has to work each individual query. Lastly, the tasks completed criterion indicates how many individual tasks or items were completed in the simulation (there are 9 items or tasks in each query). The descriptive statistics for each completion criterion variable shows adequate variance and the pattern of relationships make conceptual sense (see Table 6). The completion criteria are strongly correlated with one another and with the total accuracy criterion. (Note: the total accuracy criterion and the tasks attempted criterion are extremely highly correlated, which makes sense because the "accuracy attempted" descriptive statistics show that when items are completed, they are typically done so accurately. In other words, the tasks attempted and total accuracy criteria are tapping largely the same construct).

Speed Criterion. There are two timed measures collected for each query: how long it took for the participant to open the query and how long it took the participant to submit the query. On the surface, how long it takes the participant to open a query appears to be a reaction time measure, but the reality is more complicated. Because the participants are working on other tasks when most of the queries arrive, participants may notice the new query and choose not to open it until they finish the previous query. Therefore, the reaction time is confounded by performance on the previous query and the strategies adopted by the participant in the simulation. Also, because there is a great deal of variance in the number of queries opened by each participant (completion criterion), comparisons of average reaction times are difficult to interpret.

Likewise the average time it takes to submit a query depends on the completion and accuracy strategies adopted by the participant. If the participant opts to complete as many tasks as possible with no concern for accuracy, these times are likely to be short. If the participant is more concerned about accuracy, these times are likely to be long.

Despite these concerns, however, the speed criterion variables do have variance and the pattern of relationships with the other criterion variables is as would be expected (see Table 6). The speed variables have a moderate positive relationship with total accuracy, but it has a moderate negative relationship with the completion criteria. This pattern of results means that people who take longer to open queries and complete tasks do better on the simulation. This finding is not surprising. The tasks involved in each query take several seconds to complete. The participants must use most of the allotted sixty seconds to complete each query accurately, which also means that queries in the queue often have to wait several seconds before they are opened. The negative relationship with the completion criteria means that the longer participants take to complete tasks, the fewer they are able to complete. Again, this finding should be expected. In the end, the speed variables are worth considering, but will likely take a back seat to the total accuracy and completion criteria in evaluating the hypotheses due to the difficulty in interpreting them given their lack of independence from the other criteria.

A confirmatory factor analysis was conducted to examine the structure of the criterion variables. This analysis was used to test the hypothesis that the three accuracy and completion criteria (total accuracy, queries opened and queries submitted) and the two speed criterion (open time and submit time) could be grouped into two broad factors (accuracy/completion and speed). The two-factor model (with correlated factors) fit the data well (RMSEA = .03).

RESULTS

Task Characteristic Hypotheses

The task characteristic hypotheses are within-subjects predictions. That is, all three hypotheses predict that individual performance will vary at different points in the multitasking simulation based on the levels of different task characteristics. These within-subjects predictions are best examined using repeated measures ANOVA procedures. This approach identifies performance variance under different task characteristic situations in terms of mean differences on the dependent variable.

While five potential criterion variables were discussed in the methods section, only the accuracy criterion makes sense as a dependent variable in considering the task characteristic hypotheses. This is the only criterion variable on which meaningful comparisons at different points in the multitasking simulation can be made. Fortunately, this operationalization of performance in a multitasking environment is also the operationalization that makes the most sense conceptually as a measure of overall performance in the multitasking simulation (as discussed in the methods section).

Before these hypotheses were tested, the descriptive statistics were analyzed for each individual query. The descriptive statistics for each individual query are important because these hypotheses are based on various task characteristics resulting from an increasing multitasking demand manipulation for each successive query. These descriptive statistic results are reported in Table 7.

Table 7

Multitasking Simulation Descriptive Statistics from the Current Sample (n = 2711 Job Candidates)

Trial #	Query #	Delivery Time	Min.	Max.	Mean	SD
Trial 1	Query 1	Start	.00	1.00	.55	.39
	Query 2	55 sec. Later	.00	1.00	.48	.33
	Query 3	50 sec. Later	.00	1.00	.45	.32
	Query 4	45 sec. Later	.00	1.00	.50	.38
	Query 5	40 sec. Later	.00	1.00	.38	.36
	Query 6	35 sec. Later	.00	1.00	.28	.30
	Query 7	40 sec. Later	.00	1.00	.42	.40
	Query 8	40 sec. Later	.00	1.00	.39	.32
	Total	Total Trial Score	.00	1.00	.43	.26
Trial 2	Query 1	Start	.00	1.00	.57	.35
	Query 2	55 sec. Later	.00	1.00	.59	.34
	Query 3	50 sec. Later	.00	1.00	.63	.38
	Query 4	45 sec. Later	.00	1.00	.48	.32
	Query 5	40 sec. Later	.00	1.00	.46	.39
	Query 6	35 sec. Later	.00	1.00	.38	.39
	Query 7	40 sec. Later	.00	1.00	.45	.40
	Query 8	40 sec. Later	.00	1.00	.50	.34
	Total	Total Trial Score	.00	1.00	.51	.28

In comparing the means reported in Table 1, which come from the initial pilot of the multitasking simulation conducted with job incumbents, with the means reported in Table 7, which comes from a sample of job candidates, a few notable points emerge. This comparison is represented in a line graph in Figure 5. First, the means from the incumbent sample are significantly higher than the means from the candidate sample. This mean difference was expected. The sample of job incumbents used in the pilot is much more range-restricted than the sample of job candidates. The job incumbents were screened based on additional selection tools that had not yet been used with the candidate sample. Further, the incumbent sample consists of participants that had been working in full time jobs with demanding multitasking demands for a minimum of three months. No such consistent multitasking experience existed in the candidate sample. Therefore, it is plausible to expect that the means in an incumbent sample would be higher than the means in the candidate sample (in fact, if this were not the case, the utility of the host organization's selection system would be fair to question).

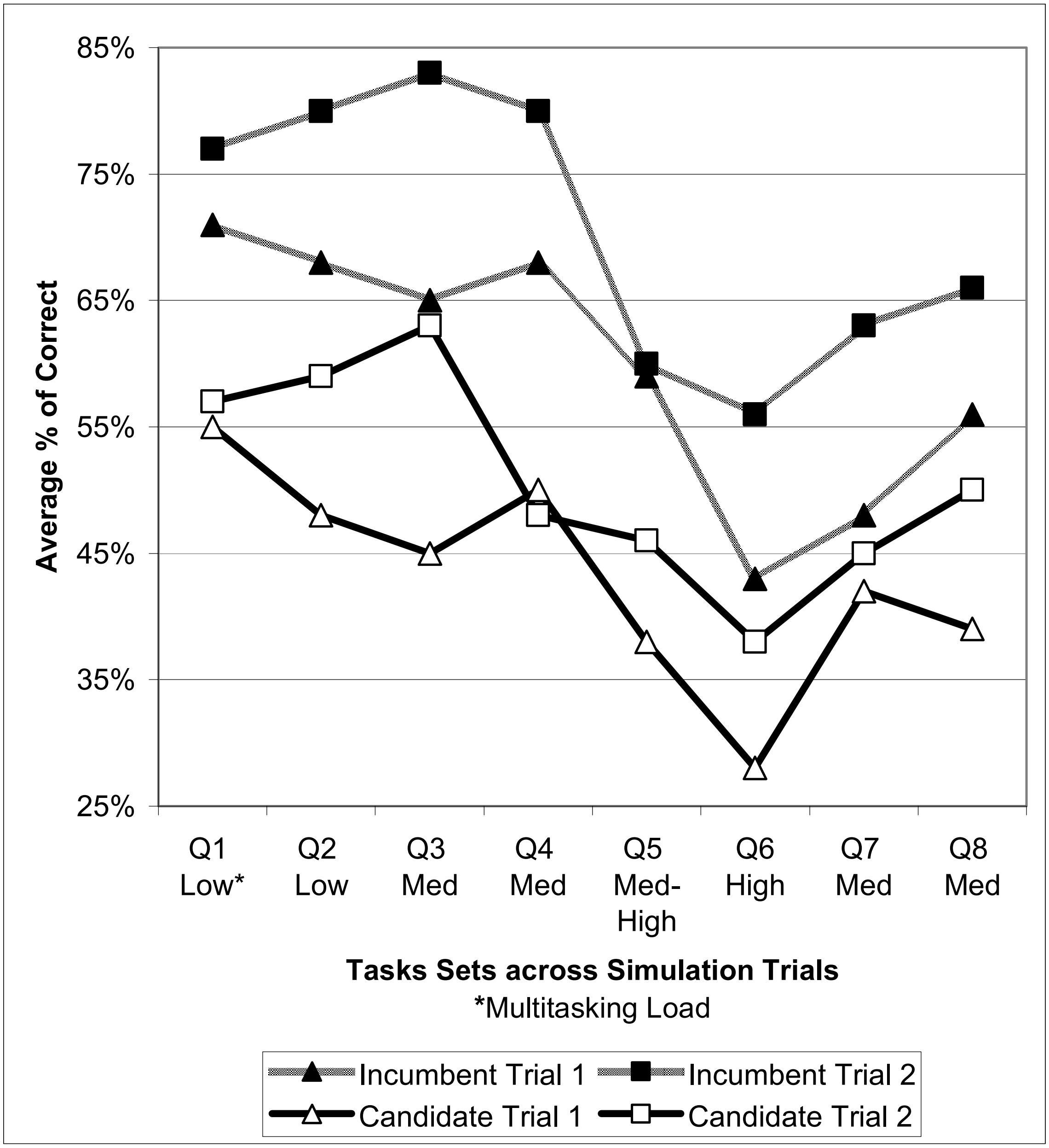

Figure 5. Patterns of Performance Across Trial One and Trial Two in the Incumbent and Candidate Samples.

The second point of note from a comparison of Table 1 and Table 7 is the direction of mean changes across the simulation trials. While the magnitude of the means is very different, the impact on performance of the multitasking demand manipulation is remarkably similar. In other words, the direction of mean change when moving from one query to the next is the same for every possible comparison across both trials. This trend is illustrated for trial one in Figure 6, and it is illustrated for trial two in Figure 7. This finding suggests that the multitasking manipulation had the same general impact on both samples.

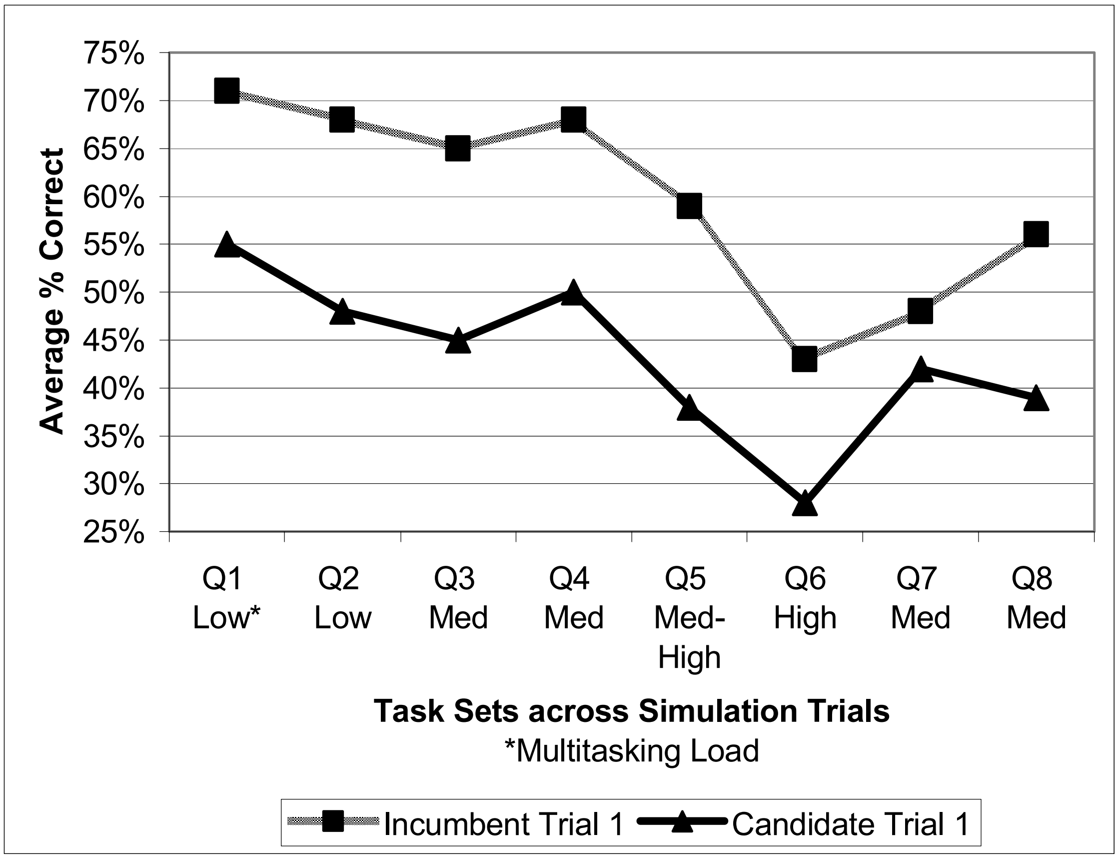

Figure 6. Patterns of Performance in Trial One in the Incumbent and Candidate Samples.

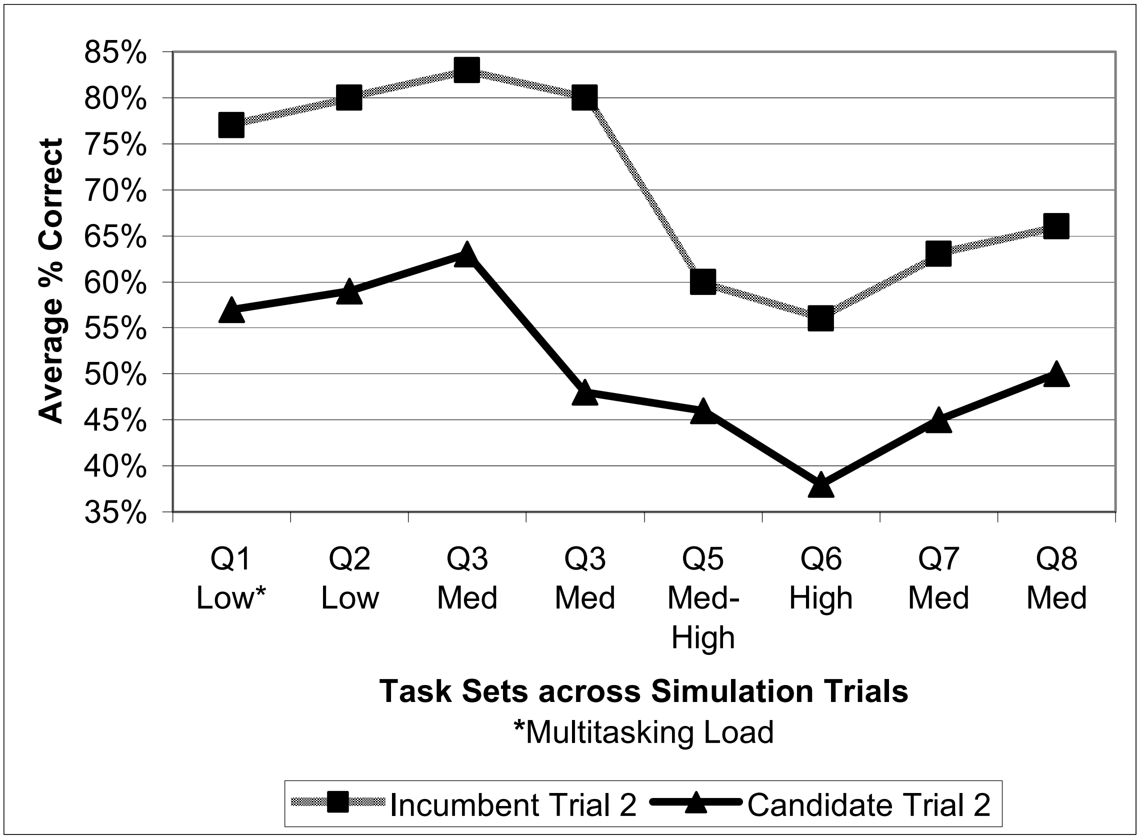

Figure 7. Patterns of Performance in Trial Two in the Incumbent and Candidate Samples.

The final point of note from a comparison between Table 1 and Table 7 is the similarity of the standard deviations for each query. Specifically, the standard deviations for each of the 16 possible comparisons across Table 1 and Table 7 are within a magnitude of .06 with the exception of one comparison (query three in trial two SD = .11). While it would be expected that the means would change based on the nature of the two samples, the standard deviations should remain relatively consistent. If they did not, then the reliability of the measures in the simulation could be called into question, and it could be questioned whether or not the observed variance was due to random error. The consistent pattern of the standard deviations suggests that the observed variance is predictable and likely due to the consistent multitasking manipulation.

Regarding the standard deviations in both samples, it is also important to note that the standard deviations stay relatively stable across queries in the simulation (see Tables 1 and 7). This is important because it lends still more evidence of reliability to the multitasking simulation. If it were the case that the standard deviations changed drastically across different task characteristics, the meaning of any observed differences could be called into question because the observed variance could have been due to random error rather than the task characteristic under investigation.

Based on the descriptive statistics analysis, it was determined that meaningful conclusions could be reached by testing the task characteristic hypotheses using a repeated measures ANOVA approach. Therefore, to examine each of the three tasks characteristic hypotheses, a series of repeated measures ANOVA procedures were employed to understand whether or not critical task characteristics in the multitasking environment impacted the observed performance.

Hypothesis one predicted a decrease in performance in a high-demand multitasking environment relative to a monotask or low demand multitasking environment. In the simulation, query one in both trials represents a very low demand multitasking environment. Then the demand increases through query six. The demand eases up in queries seven and eight (although the situation is still much more demanding than in query 1). Therefore, the critical comparison to test this hypothesis is query one versus query six; this comparison represents the essence of this hypothesis. Further, this hypothesis would suggest that any increase in multitasking demand should lead to a decrement in performance, therefore, it was also expected that mean differences would be observed between query one and all other query comparisons. The results of this series of ANOVA procedures are reported in Table 8.

Table 8

Tests of Hypothesis One: Repeated Measures ANOVA Results

Trial	Query Comparison	d.f.	Mean Square	F-Statistic	Significance	Partial Eta Squared
Trial 1	Q1 and Q2	1	6.28	170.50	.00	.06
	Q1 and Q3	1	13.57	326.25	.00	.11
	Q1 and Q4	1	3.05	72.98	.00	.03
	Q1 and Q5	1	38.89	600.48	.00	.18
	Q1 and Q6*	1	**95.97**	**1132.67**	**.00**	**.30**
	Q1 and Q7	1	22.35	244.57	.00	.08
	Q1 and Q8	1	32.07	529.65	.00	.16
Trial 2	Q1 and Q2	1	.29	12.22	.00	.00
	Q1 and Q3	1	4.72	107.67	.00	.04
	Q1 and Q4	1	12.02	352.11	.00	.12
	Q1 and Q5	1	16.90	322.77	.00	.11
	Q1 and Q6*	1	**52.16**	**735.10**	**.00**	**.21**
	Q1 and Q7	1	21.42	259.97	.00	.09
	Q1 and Q8	1	8.24	153.33	.00	.05

Note. Comparisons between Query 1 and Query 6 represent the greatest difference in multitasking demand. Consequently, this comparison is the critical test of hypothesis one.

The results in Table 8 suggest that hypothesis one was supported. The mean difference comparisons between query one and all other queries were statistically significant (however the mean difference for query two and query three in trial two were not in the predicted direction; Figure 8 shows a graphical illustration of these mean changes). Most importantly, the comparisons between query one and query six in the two trials received the largest F-Statistics out of all of the comparisons.

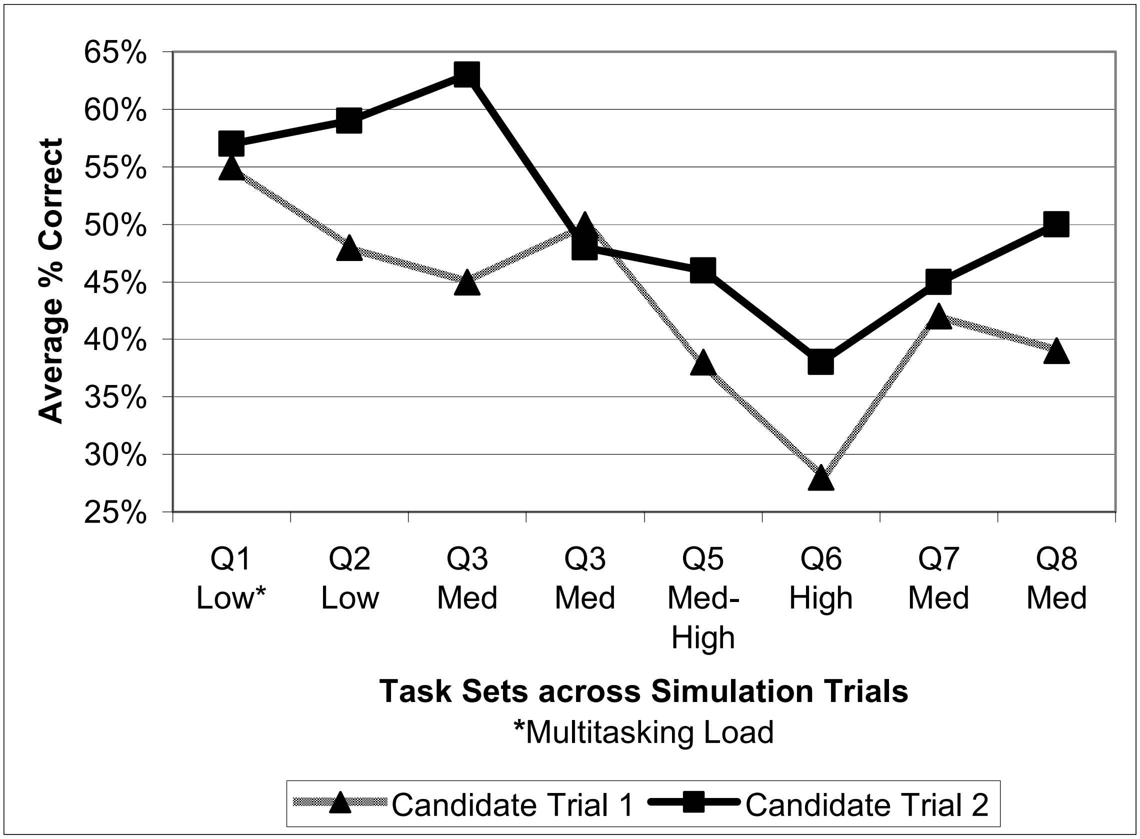

Figure 8. Patterns of Performance in Trial One and Trial Two in the Candidate Sample.

With a sample size of 2711, it is not surprising that any observed mean difference was statistically significant, so to interpret these results, an estimate of effect size is also reported. Partial-Eta^2 is calculated by dividing the sum of squares for the effect by the sum of squares of the effect plus the sum of squares for error. The result is an indication of the proportion of the variance that is due to the effect. In other words, partial-eta^2 gives an indication of the practical significance of the effect.

From examining the effect size estimates reported in Table 8, it is clear that the mean difference between query one and query six is an important mean difference (partial-eta^2 = .30 and .21 in trial one and trial two, respectively). These results suggest that the impact on performance of a high-demand multitasking environment relative to a low-demand multitasking environment is substantial. Consequently, it was determined that hypothesis one received support.

Hypothesis two predicted that having the opportunity to practice tasks would lessen the impact on performance in a multitasking environment. Specifically, this hypothesis predicted that practice would reduce, but not eliminate the multitasking demand effect. In order to test this hypothesis, comparisons between performance on trial one versus performance on trial two were examined. Because the two trials consist of parallel tasks and precisely the same manipulation, the only substantive difference between the two trials is that trial one occurs with no prior practice and trial two occurs after having practiced performing in a multitasking environment. The support received by hypothesis one and reported in Table 8 convincingly show that the effect of the multitasking demand is not eliminated in trial two. Consequently, the test of this hypothesis centers on whether or not the effect is reduced. Operationally then, this hypothesis is supported if performance on trial two queries is significantly better than performance on the parallel trial one queries. If the trial two means are higher, then practice has improved performance in a multitasking environment. Results of the comparisons between the trial one and trial two are reported in Table 9 and shown graphically in Figure 8.

Table 9

Tests of Hypothesis Two: Repeated Measures ANOVA Results

Across Trial Comparison	d.f.	Mean Square	F-Statistic	Significance	Partial Eta Squared
Query 1	1	.88	18.7	.00	.01
Query 2	1	15.82	548.01	.00	.17
Query 3	1	46.15	956.58	.00	.26
Query 4	1	.62	14.98	.00	.01
Query 5	1	9.38	163.53	.00	.06
Query 6	1	12.33	156.56	.00	.06
Query 7	1	1.08	9.16	.00	.00
Query 8	1	13.91	220.45	.00	.08
Total Score	1	7.75	1051.56	.00	.28

The results in Table 9 and Figure 8 have several important considerations relative to the practice effect hypothesis. First, all comparisons between means from trial one and trial two are statistically significant due to the large sample. Again, consideration of the partial-eta^2 effect size measure should be the standard by which hypothesis support should be evaluated. Second, the mean difference for the total score between trial one and trial two are significantly different with a partial-eta^2 of .28. This finding suggests support of this hypothesis; the means for the queries in trial two were significantly and meaningfully higher than they were for trial one.

When comparing the individual queries, support for this hypothesis does change slightly. There is no meaningful difference on query one. Because this query is a very low-multitasking demand, this finding is not critical in consideration of this hypothesis. For queries two and three very supportive results were observed. The partial-eta^2 for these two comparisons was .17 and .26, respectively. It is clear that at this level of multitasking demand, practice has improved performance. After query three, partial-eta^2 drops drastically. For queries five, six, and eight, practice does explain an important proportion of variance, but not as much as was explained in trials two and three. For trials four and seven, practice explains a practically insignificant proportion of variance. The pattern of results does suggest support for the practice hypothesis, but the relative changes in the effect across the multitasking demand (see Figure 8) does suggest that this effect depends on the intensity of the multitasking demand. This idea has practical implications for employee selection and job design, which are expanded upon in the discussion.

Hypothesis three predicted that a decreasing time interval between the response on a task and the delivery of the subsequent task would hurt performance. Operationally, this hypothesis predicts that performance declines as the delivery interval for each individual query reduces (i.e., the delivery frequency increases). To test this hypothesis, the comparisons on each successive query pair must be examined. For the hypothesis to receive support, significant mean declines in performance should be observed out to query six and then performance should improve on query seven and hold steady on query eight (this pattern is expected due to the delivery interval pattern of the queries reported in Table 7). The results of these comparisons are reported in Table 10.

Table 10

Tests of Hypothesis Three: Repeated Measures ANOVA Results

Trial	Query Comparison	d.f.	Mean Square	F-Statistic	Significance	Partial Eta Squared
Trial 1	Q1 and Q2	1	6.28	170.50	.00	.06
	Q2 and Q3	1	1.39	52.80	.00	.02
	Q3 and Q4	1	3.75	97.90	.00	.04
	Q4 and Q5	1	20.17	334.23	.00	.11
	Q5 and Q6	1	12.67	154.90	.00	.05
	Q6 and Q7	1	25.69	203.69	.00	.07
	Q7 and Q8	1	.33	6.29	.01	.00
Trial 2	Q1 and Q2	1	.29	12.22	.00	.00
	Q2 and Q3	1	2.69	68.96	.00	.03
	Q3 and Q4	1	31.82	704.79	.00	.21
	Q4 and Q5	1	.41	8.67	.00	.00
	Q5 and Q6	1	9.68	106.56	.00	.04
	Q6 and Q7	1	6.73	45.71	.00	.02
	Q7 and Q8	1	3.09	33.82	.00	.01

The results reported in Table 10 suggest general support for the hypotheses, but the effect does not appear to be as large as the effects observed in the tests of hypothesis one and two. Further there are a few particular comparisons and trends in this table that warrant consideration when evaluating this hypothesis. First, given that the tests of hypotheses one and two show that the nature of performance in trial two is somewhat different than trial one, the two trials are considered separately as unique tests of hypothesis three.

In trial one, the first two comparisons (queries one and two; queries two and three) are in the predicted direction and statistically significant, although the practical impact of the effect is relatively small (mean changes across trial one are illustrated in Figure 9). Then, in comparison three (query 3 and 4), the direction of the mean difference reverses. Therefore, even though the delivery interval reduces, the performance increases. This pattern of results suggests that, in trial one, when the multitasking demand is relatively low, within-trial practice (i.e., the practice effect of doing subsequent queries) accounts for more variance than the variance accounted for by the reducing delivery interval. Then, on the fourth comparison (query four and five), the participants "hit the wall". That is, the increased demand from the reducing delivery interval catches up with and overcomes the variance accounted for by the within-trial practice effect. This explains why such a large partial-eta^2 relative to the others in this trial is observed on this comparison. In essence, the effect size represents the cumulative impact of the reducing interval on the first four comparisons. After this spike on the fourth comparison, the effects are in the predicted direction and relatively consistent for the remaining comparisons. In comparison five (query five and six) the partial-eta^2 is not large, but there is a practically significant decline in performance. In comparison six (query six and seven), performance improves to a small but still practically significant degree. This improvement is expected because the delivery interval increases (rather than reduces) for this comparison. Finally, in the final comparison of the trial (query seven and eight), there is no effect on performance. Again, this is expected because the delivery interval remains static for these two queries.

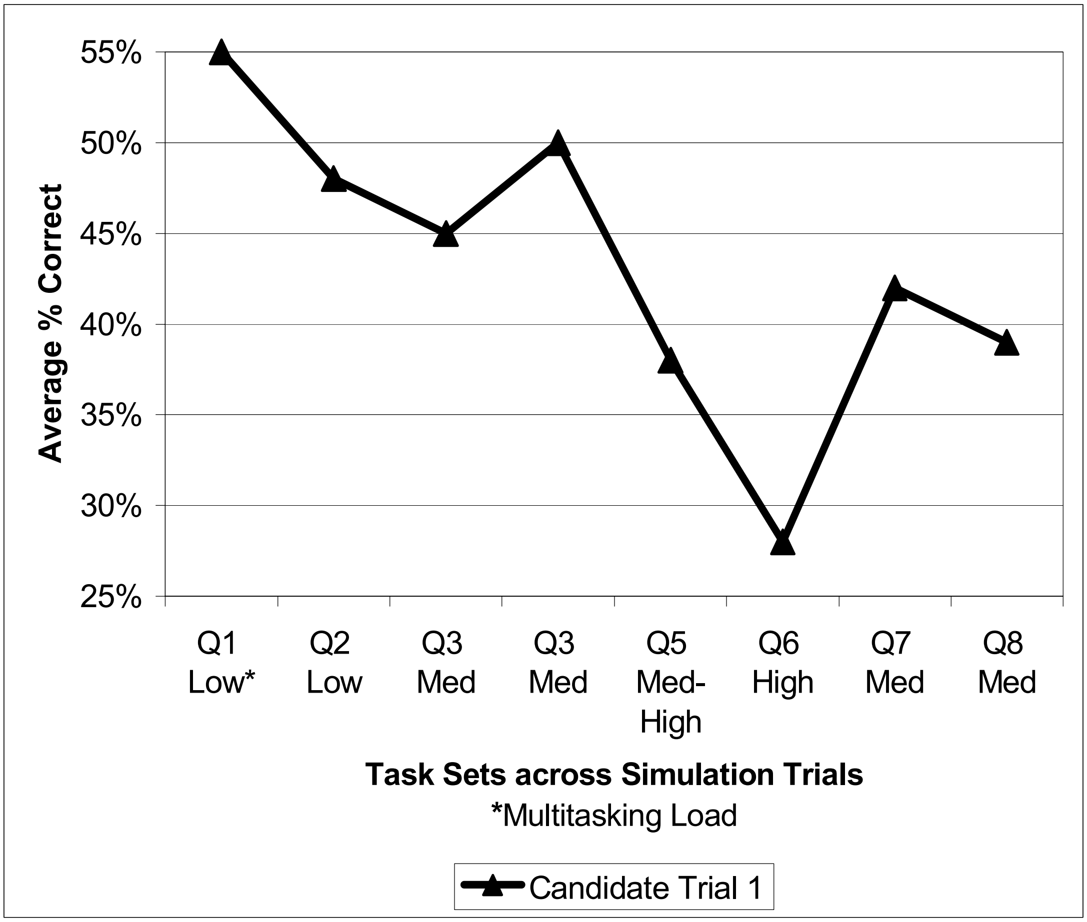

Figure 9. Pattern of Performance in Trial One in the Candidate Sample.

In summary, for trial one hypothesis three is supported. In the first half of the trial, the response-stimulus effect does not take hold until the multitasking demand becomes suitably intense, but when it does, a large impact on performance is observed. The hypothesis is supported as predicted for the second half of the trial where the multitasking demand is more intense. Consequently, in trial one hypothesis three was supported.

In trial two, a similar dynamic occurred, although the pattern is slightly different than in trial one (mean changes across trial two are illustrated in Figure 10). In trial two, performance increases across the first two comparisons. It is likely the case that the trial level practice effect observed in the test for hypothesis two combined with the within-trial practice effect trump the impact of the reducing delivery time in very low multitasking demand situations.

Interestingly, however, the participants "hit the wall" earlier in trial two than in trial one. On the third comparison (queries three and four), a drastic drop in performance is observed with a large effect size (partial-eta^2 = .21). Then on comparison four (queries four and five) there is no significant change in performance. This pattern of results suggest that with practice a very low multitasking demand will not impact performance, but when the demand increases slightly performance drops significantly. However, if the multitasking demand stays at this mid-level, with practice, perhaps the reduced performance level can be maintained.

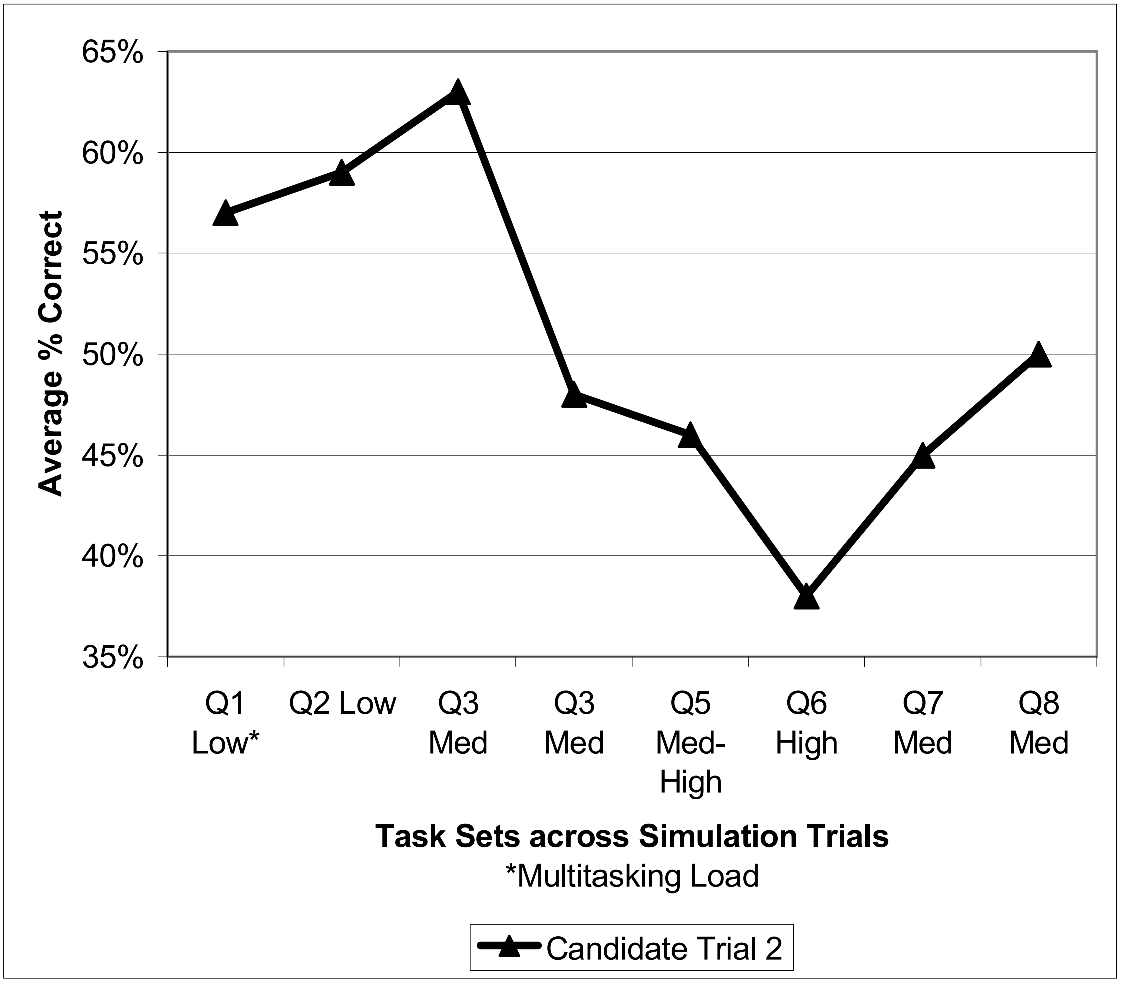

Figure 10. Pattern of Performance in Trial Two in the Candidate Sample.

Once the multitasking demand increases beyond the mid-level, the stimulus-response hypothesis holds true. As was the case in trial one, after the multitasking demand becomes intense in comparison five (query five and six), the impact on performance moves in the predicted direction for the remainder of the trail. In other words, the performance goes down for comparison five, up for comparison six, and there is relatively no change in performance for comparison seven.

To summarize the test of hypothesis three in trial two, practice effects do trump the response-stimulus hypothesis under low multitasking demand conditions. However, as was seen in trial one, once the multitasking demand reaches a certain threshold, the stimulus-response interval hypothesis is supported. Additionally, the test of this hypothesis identifies the impact on performance that low-, medium-, and high-multitasking demand situations have on participants. The different effects at the three different levels have important implications for employee selection and job design. These implications will be explained in the discussion. Table 11 outlines hypothesis support for the task characteristic hypotheses.

Table 11

Support Summary for the Task Characteristics Hypotheses

#	Hypothesis	Was the Hypothesis Supported?
1	An adverse effect on performance relative to a single task environment will occur when attentional resources must be shared across multiple tasks.	Yes
2	Practice on the simulation tasks will reduce, but not eliminate the performance decrement created by the multitasking environment	Yes
3	As the time interval between the response to the previous task and the delivery of the next task decreases, task performance will also decrease.	Yes

Individual Difference Hypotheses

Summary descriptive statistics were examined to ensure that the measures used in this study received similar results as were observed in past studies. First, the descriptive statistics and intercorrelation coefficients for the predictor measures were examined. These statistics are reported in Table 12. (Note that the descriptive statistics for the criterion variables were discussed in the methods section and reported in Table 6).

Table 12

Descriptive Statistics, Inter-Correlation Coefficients, and Reliability Coefficients from the Current Sample (n = 2711)

Predictor Measures	Mean	SD	1	2	3	4	5	6	7	8	9
1 Basic Skills Test	22.50	7.02	*.90*								
2 Extraversion	3.69	.45	.11	*.91*							
3 Intellectance	4.32	.39	.17	.56	*.79*						
4 Conscientiousness	4.31	.38	-.15	.40	.54	*.83*					
5 Agreeableness	4.26	.37	.04	.54	.59	.54	*.83*				
6 Stress Tolerance	3.97	.56	.04	.63	.60	.54	.55	*.74*			
7 Competitiveness	2.82	.76	.04	.29	.11	.07	-.06	.13	*.76*		
8 Ach. Striving	4.22	.44	-.10	.47	.50	.75	.40	.44	.22	*.63*	
9 Polychronicity	4.00	.55	.14	.40	.46	.38	.43	.53	.06	.28	*.60*

Notes. Reliability coefficients are reported along the diagonal in italicized font. All coefficients > r = .05 are significant at p < .05. BST reliability taken from a different sample; item-level data was not included in the data set. The reliability for the BST is from a sample of 421 with a mean of 22.37 and a standard deviation of 7.73.

The descriptive statistics suggest that there is some range restriction in a few of the predictors (e.g., intellectance: mean = 4.32, SD = .39), but even in these cases there was still adequate variance such that relationships should not be suppressed. The intercorrelation coefficients observed in this sample were parallel to those observed in past settings. In fact, the results in Table 12 match the results from the past studies reported in Table 4 within a magnitude of .04. Further, the reliability coefficients for the predictors on the Personal Preferences Inventory (PPI) are parallel to the reliability coefficients observed in past research. Comparing Table 5 and Table 12 shows that all of the scales' internal consistency reliability varies less than a magnitude of .03, with the exception of the achievement striving scale which only varied with a magnitude of .05.

For further evidence of construct validity, a confirmatory factor analysis was conducted using the PPI data to see if a model with all of the predictors loading onto different factors would fit the data. The results of the confirmatory factor analysis suggest that this model fit the PPI data well (RMSEA = .05). Other alternative models were also investigated to see if better fit could be achieved. For example, a model using only the predictors with high levels of reliability was investigated, but the fit was not significantly better (RMSEA = .05). The results of the confirmatory factor analysis add further evidence of the construct validity of the predictor tool.

The internal consistency reliability for the cognitive ability test could not be computed, because the item-level data was not available in this dataset. Therefore, reliability results from a different data set that included item-level data are reported in Table 12. It is appropriate to generalize this reliability coefficient given the following considerations. (a) The pattern of predictor intercorrelation coefficients is virtually the same in this study as has been observed in past studies. (b) The other predictors' reliability coefficients have remained stable over time (thus unreliability driven by participant differences is unlikely to occur). (c) The descriptive statistics in the adequately large sample (n=421) in the past data set are very similar to those observed in the current sample (Current sample mean and SD = 22.50 and 7.02, respectively; Earlier sample mean and SD = 22.37 and 7.73, respectively).

In order to examine the relationship between the individual difference predictors and performance in a multitasking environment correlation coefficients were generated for each of the proposed predictor/criterion relationships. As discussed in the methods section, there are five potential criterion variables that warrant consideration from the multitasking simulation, however, the most important criterion measure is the accuracy criterion, as it is the most complete and most interpretable operationalization of overall performance in a multitasking environment. Therefore, the prediction of this variable is the standard by which support of the hypotheses was judged. Correlation coefficients representing tests of each of the nine individual difference hypotheses are reported in Table 13.

Table 13

Tests of the Individual Difference Hypotheses: Predictor/Criterion Validity Coefficients

Criteria	BST	Ext	Int	Con	Agr	ST	Comp	AchStr	PolyC
Accuracy	.67*	.12*	.20*	-.097*	.03	.04*	.05*	-.03	.13*
Queries Opened	.33*	.08*	.11*	-.02	.02	.04*	.04*	.00	.06*
Queries Submitted	.41*	.06*	.10*	-.08*	-.01	.01	.02	-.03	.05*
Avg. Open Time**	.29*	.02	.06*	-.08*	-.04	.02	-.01	-.05	.08*
Avg. Submit Time**	.43*	.10*	.14*	-.08*	.01	.05	.02	-.02	.12*

Notes. * = Significant at p-value <.05. ** = The correlation coefficients for reaction time and completion time were taken from a subset of the sample of people that completed at least two thirds of the 16 queries. This approach was chosen to ensure stability in the average speed variables. BST = Basic Skills Test (cognitive ability); Ext = Extroversion; Int = Intellectance; Con = Conscientiousness; Agr = Agreeableness; ST = Stress Tolerance; Comp = Competitiveness; AchStr = Achievement Striving; PolyC = Polychronicity.

The first individual difference hypothesis (hypothesis four), predicted that cognitive ability is a strong predictor of performance in a multitasking environment. The correlation between the cognitive ability test and the accuracy criterion measure was .67, which constitutes a strong relationship. Thus, the cognitive ability hypothesis was supported. This hypothesis was also partially supported for all of the alternative criterion operationalizations. The correlation magnitudes for these other variables ranged from .29 to .43. While these relationships are only moderate in size, it is clear that cognitive ability is an influential predictor of performance in a multitasking environment.

Hypothesis five predicted a weak relationship between extroversion and performance in a multitasking environment. The correlation between extroversion and the accuracy criterion was .12. Thus, the extraversion hypothesis was supported. This hypothesis was also supported for the average submit time criterion variable where a correlation of .10 was observed, however, it was not supported for the remaining three criterion variables. The magnitude of the relationship between extraversion and the three remaining criterion variables was under .10, which is not practically important.

Hypothesis six predicted a moderate relationship between intellectance and performance in multitasking environment. This hypothesis received only partial support. The correlation between intellectance and the accuracy criterion was .20. While practically and statistically significant, this relationship did not meet the .30 threshold to be considerate moderate. Rather, intellectance appears to be a weak predictor of performance in multitasking environment. Even though the magnitude of this relationship is not as large as predicted, because this variable did emerge as a significant predictor of the accuracy criterion, the hypothesis did receive partial support. The finding that intellectance is a weak predictor of performance in multitasking environment was consistent across three of the four alternative criterion operationalizations. The magnitudes of these relationships ranged from .10 to .14. One of the five criterion variables was not predicted with a practically significant magnitude by intellectance.

Hypothesis seven predicted that conscientiousness would not predict performance in a multitasking environment. The relationship between conscientiousness and the accuracy criterion was -.097. While statistically significant, this relationship falls below the .10 threshold of practical significance adopted in this study. Consequently, this hypothesis is supported. Further, the relationships between conscientiousness and the five alternative criterion measures were also not practically significant. The magnitudes for these relationships ranged from -.02 to -.08.

Hypothesis eight predicted that agreeableness would not predict performance in a multitasking environment. The relationship between agreeableness and the accuracy criterion was .03, which is neither practically significant nor statistically significant. Thus, it is safe to conclude that there is no relationship between agreeableness and performance in a multitasking environment in this data set. In further support, the relationships between agreeableness and all of the alternative criterion variables did not meet the threshold for either practical significance or statistical significance.

Hypothesis nine predicted that stress tolerance would be a weak predictor of performance in a multitasking environment. This hypothesis was not supported. The relationship between stress tolerance and the accuracy criterion was .04, which does not meet the .10 threshold for a weak relation. Stress tolerance did not predict the remaining five criterion operationalizations. Consequently, there is no relationship between stress tolerance and performance in a multitasking environment according to this data set.

It should be noted, however, that the lack of relation between stress tolerance and the criterion variables could be due to the low reliability of the predictor variable. The internal consistency reliability coefficient for stress tolerance was only .74, which is below the conventional .80 standard. This lower reliability could have attenuated the predictor-criterion relationships.

Hypothesis ten predicted that competitiveness would be a moderate predictor of performance in a multitasking environment. The relationship between competitiveness and accuracy was only .05. The hypothesis predicted that the relationship would be at least .30. Consequently, this hypothesis is not supported. Further, there were no practically significant relationships observed between competitiveness and any of the alternative predictors.

Just like the stress tolerance scale, the competitiveness has a lower than ideal reliability coefficient. The internal consistency for competitiveness was only .76. This lower than convention reliability coefficient may have attenuated the predictor-criterion relationship somewhat, however, even if this attenuation was corrected for, the hypothesis would still not be supported. In this case, regardless of the reliability issues in the predictor, the hypothesis did not hold up.

Hypothesis 11 predicted that achievement striving would be a moderate predictor of performance in a multitasking environment. The relationship between achievement striving and the accuracy criterion was -.03, which is not statistically significant or practically significant. Thus, this hypothesis did not receive support. Further the relations between achievement striving and the alternative criteria also were not statistically significant or practically significant.

The reliability coefficient for the achievement striving scale is very low at .63. With such a low reliability, it is possible that meaningful relationships would emerge between achievement striving and performance in a multitasking environment if a better measure of achievement striving were used. Therefore, while the hypothesis was not supported in this study, it warrants further investigation to conclusively disconfirm that achievement striving is a moderate predictor of performance in a multitasking environment.

Finally, hypothesis 12 predicted that polychronicity would be a weak predictor of performance in a multitasking environment. The relationship between polychronicity and the accuracy criterion was .13, which meets the threshold for a weak relationship. Consequently, this hypothesis did receive support. Further, this relationship was also supported for the "average submit time" criterion variable where a correlation coefficient of .12 was observed. The remaining three alternative criteria were statistically, but not practically significant.

The polychronicity scale also suffered from reliability problems. The root of the reliability problems for this scale likely lies in the fact that it consists of only four items. Therefore, it is more difficult to reach the .80 conventional standard for internal consistency reliability. Given that the polychronicity scale was strongly correlated with a more comprehensive measure of polychronicity in other research (as discussed in the methods section), it is appropriate to conclude that the observed correlation between polychronicity and the accuracy criterion in this study reflects a meaningful relation between the two constructs.

In sum, five of the nine individual difference hypotheses were supported, with one more receiving partial support. Three of the individual difference hypotheses were not supported. Support for all of these hypotheses is listed in Table 14.

Table 14

Support Summary for the Individual Difference Hypotheses

#	Hypothesis	Was this Hypothesis Supported?
4	A psychometric cognitive ability test (**reading comprehension**) is a **strong predictor** of performance in a multitasking environment.	Yes
5	**Polychronicity** is a **weak predictor** of performance in a multitasking environment.	Yes
6	**Achievement Striving** is a **moderate predictor** of performance in a multitasking environment.	No
7	**Competitiveness** is a **moderate predictor** of performance in a multitasking environment.	No
8	**Extraversion** is a **weak to moderate predictor** of performance in a multitasking environment.	Yes
9	**Stress tolerance** is a **weak predictor** of performance in a multitasking environment.	No
10	**Intellectance** is a **moderate predictor** of performance in a multitasking environment.	Partially
11	**Conscientiousness does not predict** performance in a multitasking environment.	Yes
12	**Agreeableness does not predict** performance in a multitasking environment.	Yes

From the correlation analyses, four variables emerged as important predictors of performance in this multitasking environment. These predictors were cognitive ability, extroversion, intellectance, and polychronicity. To understand the relative importance of each of these predictors, a series of hierarchical regression analyses were completed. These analyses investigated changes in explained variance as different predictors and combinations of predictors are added to and removed from the predictor model. The results of these analyses are reported in Table 15.

Table 15

Hierarchical Regression Results for Significant Predictors Across All Predicted Criteria

Criterion	Model		R	R^2	ΔR^2
Total Accuracy	Step 1:	1	.67	.45	--
	Step 2:	1, 2	.68	.46	.01
	Step 3:	1, 2, 3	.68	.46	.00
	Step 4:	1, 2, 3, 4	.68	.46	.00
	Step 1:	1	.67	.45	--
	Step 2:	1, 2, 3, 4	.68	.46	.01
	Step 1:	2, 3, 4	.20	.04	--
	Step 2:	1, 2, 3, 4	.68	.46	.42
Average Submit Time	Step 1:	1	.43	.19	--
	Step 2:	1, 2	.44	.19	.01
	Step 3:	1, 2, 3	.44	.20	.00
	Step 4:	1, 2, 3, 4	.44	.20	.00
	Step 1:	1	.43	.19	--
	Step 2:	1, 2, 3, 4	.44	.20	.01
	Step 1:	2, 3, 4	.15	.02	--
	Step 2:	1, 2, 3, 4	.44	.20	.18
Queries Opened	Step 1:	1	.33	.11	--
	Step 2:	1, 2	.34	.11	.00
Queries Completed	Step 1:	1	.41	.17	--
	Step 2:	1, 2	.42	.17	.00

Notes. 1 = Cognitive Ability; 2 = Intellectance; 3 = Polychronicity; 4 = Extraversion; Only predictors that correlated above r = .10 with the given criterion variable were included in the hierarchical regression analyses.

Table 15 displays the results of several hierarchical regression analyses across four of the different multitasking simulation performance criteria. Based on the correlational evidence, the fifth performance criterion, average open time, was not predicted by any of the other variables except for cognitive ability. Consequently, it was not considered in these analyses. For each performance criterion, regression models were built using combinations of all of the predictors that were related to that particular criterion variable. For example, the queries completed criterion variable was only predicted by cognitive ability and intellectance. Therefore, regression models focused only on understanding the incremental variance explained by each of these predictors.

Not surprisingly, cognitive ability appears to be the major contributor to the explained variance. In fact, for all of the models across all of the criteria regardless of how they are entered, the non-cognitive predictors account for a very small change in R^2. When cognitive ability is included in the model, the regression model accounts for a large percentage of variance, but without cognitive ability very little variance is explained. This finding holds true regardless of the criterion variable used in the analysis.

The criterion variable predicted the best is the total accuracy criterion, where 46% of the total variance is explained in a model with all of the predictors included. 45% of the variance is explained by a model including only cognitive ability. Consequently, the unique variance explained by all of the remaining non-cognitive predictors is only 1% of the total variance. Of the non-cognitive predictors, intellectance is the best predictor. Without cognitive ability in the model, the non-cognitive predictors explain 4% of the total variance. This degree of prediction is important, but much less so than the large contribution from cognitive ability.

While there is some prediction from the non-cognitive predictors, the overwhelmingly skewed proportion of variance explained by cognitive ability above and beyond the non-cognitive predictors suggests that cognitive ability may be a key determinant of performance in a multitasking environment. Further, these results suggest that it may be the case that “multitasking ability” may be a sub-stratum in the hierarchy of ‘g’. This idea and the implications of such a finding are discussed in more detail in the discussion section.

Post-hoc analyses

Post-hoc analyses were performed to investigate whether or not certain demographic variables related to differences in performance in an applied multitasking environment. In particular, gender and age were chosen as potentially important demographic variables to examine. Given that cognitive ability tests often have adverse impact against female job candidates at higher cut score levels and given that performance in a multitasking simulation is highly correlated with cognitive ability, it is important to understand any potential differential effects across gender. Age was chosen because past research (e.g., Salthouse et al, 1998) has observed performance declines across age groups. Consequently, this study provides an opportunity to replicate these findings and extend them to a work-related multitasking environment.

Table 16 reports the mean differences across the demographic variables under investigation. The results in this table suggest that there are some mean differences across gender on the cognitive ability test and the multitasking simulation although these differences do not appear to be very large. Gender differences on the non-cognitive predictors are very small.

Table 16

Gender and Age Mean Differences on Predictor and Criterion Variables

Variable	Full Sample (n=1657)	Male (n=601)	Female (n=1041)	d-value (gender)	Under 40 (n=1411)	40 and Over (n=226)	d-value (age)
BST	22.70	23.41	22.26	.17	23.22	19.65	.52
Ext	3.69	3.75	3.65	.22	3.71	3.50	.47
Int	4.33	4.36	4.32	.11	4.36	4.17	.50
Con	4.33	4.32	4.34	-.05	4.35	4.20	.41
Agr	4.28	4.23	4.31	-.22	4.29	4.20	.24
ST	3.97	4.07	3.92	.27	4.00	3.86	.25
Comp	2.80	3.11	2.62	.65	2.82	2.67	.20
AchStr	4.22	4.23	4.21	.05	4.25	4.04	.48
PolyC	4.02	3.99	4.04	-.09	4.03	3.93	.19
MT Trial 1	.43	.47	.41	.23	.46	.26	.77
MT Trial 2	.51	.55	.49	.22	.54	.34	.74
MT Total	.47	.51	.45	.26	.50	.30	.87

Notes. BST = Basic Skills Test (cognitive ability); Ext = Extroversion; Int = Intellectance; Con = Conscientiousness; Agr = Agreeableness; ST = Stress Tolerance; Comp = Competitiveness; AchStr = Achievement Striving; PolyC = Polychronicity

The mean differences for age appear to be more meaningful. For the initial analyses, age groupings were created based on whether a participant was over or under the age of 40-years-old. This grouping was chosen because 40-years-old is the age at which a person becomes a member of a protected class (thus, this breakout is of particular interest to selection practitioners). The results suggest that people over 40-years-old score consistently lower on all of the measures. The age effect is lower for the non-cognitive predictors, but for the cognitive predictors the influence of age appears to be large. These results should be interpreted cautiously because of the large differences in cell size between the two groups (under 40 n = 1411; over 40 n = 226). It is also important to note that there are reasons other than multitasking differences that could account for some of the age-related differences. These possible alternative explanations are expanded in the discussion section.

To understand the effect of age on performance in a multitasking environment more fully, descriptive statistics for the multitasking simulation were examined for a more stratified grouping of the age variable. Table 17 reports the means on the multitasking simulation for participants in 10-year age groupings. Consistent with Salthouse et al. (1998), the mean performance declines as the age of the participants increases. This trend does not hold for the over 70-years-old group, but the n of 3 is too small to be meaningful. These summary statistics suggest that age may be an important factor in understanding performance in a multitasking simulation.

Table 17

Multitasking Simulation Means Differences Across Ten-Year Age Groupings

Age Group	N	MT Trial 1	MT Trial 2	MT Total
18 – 29	1092	.49	.57	.53
30 – 39	319	.37	.46	.42
40 – 49	140	.27	.34	.31
50 – 59	80	.25	.32	.28
60 – 69	3	.28	.31	.30
70 and older	3	.33	.50	.41

To examine these potential gender and age effects further, an ANOVA procedure was used to gain a greater understanding of these mean differences. Table 18 reports the results of these analyses. Specifically, these ANOVA analyses sought to discover whether or not the observed mean differences were statistically and practically significant. When gender is entered as an independent variable explaining performance on the multitasking simulation, the F-statistic is statistically significant, however the partial-eta^2 is only .01. Therefore, gender does not appear to be a practically significant predictor of performance in a multitasking environment.

Table 18

Results Testing for Significant Mean Differences on the Multitasking Simulation for Age and Gender Groups

Demographic Factor	d.f.	Mean Square	F-Statistic	Significance	Partial Eta Squared
Age Group*	3	3.77	63.83	.00	.11
Gender	1	1.15	17.58	.00	.01

Notes. * = Age groups over 60 were not included due to inadequate cell sizes. Post-hoc Bonferroni comparisons showed that all possible comparisons were significantly different except the comparison between the 40-49-years-old group and the 50-59-years-old group.

The results for the ANOVA for age in Table 18 do appear to suggest that age is an important determinant of performance. These results are shown graphically in Figure 11. The ten-year age groups were used as the independent variable in this analysis. Due to the small cell sizes, the groupings over age 60 were eliminated. The partial-eta^2 for the age effect was .11. Post-hoc comparisons were then examined to further understand the nature of this age effect. These results showed that as age increases, the means get lower. All of the possible comparisons were significant with the exception of the comparison between the 40- to 49-year-old group with the 50- to 59-year-old group. The findings of these age analyses suggest that the results from the Salthouse et al. (1998) do generalize to this applied setting. Further implications of these analyses are included in the discussion section, as well as a discussion of some other plausible explanations for this age effect.

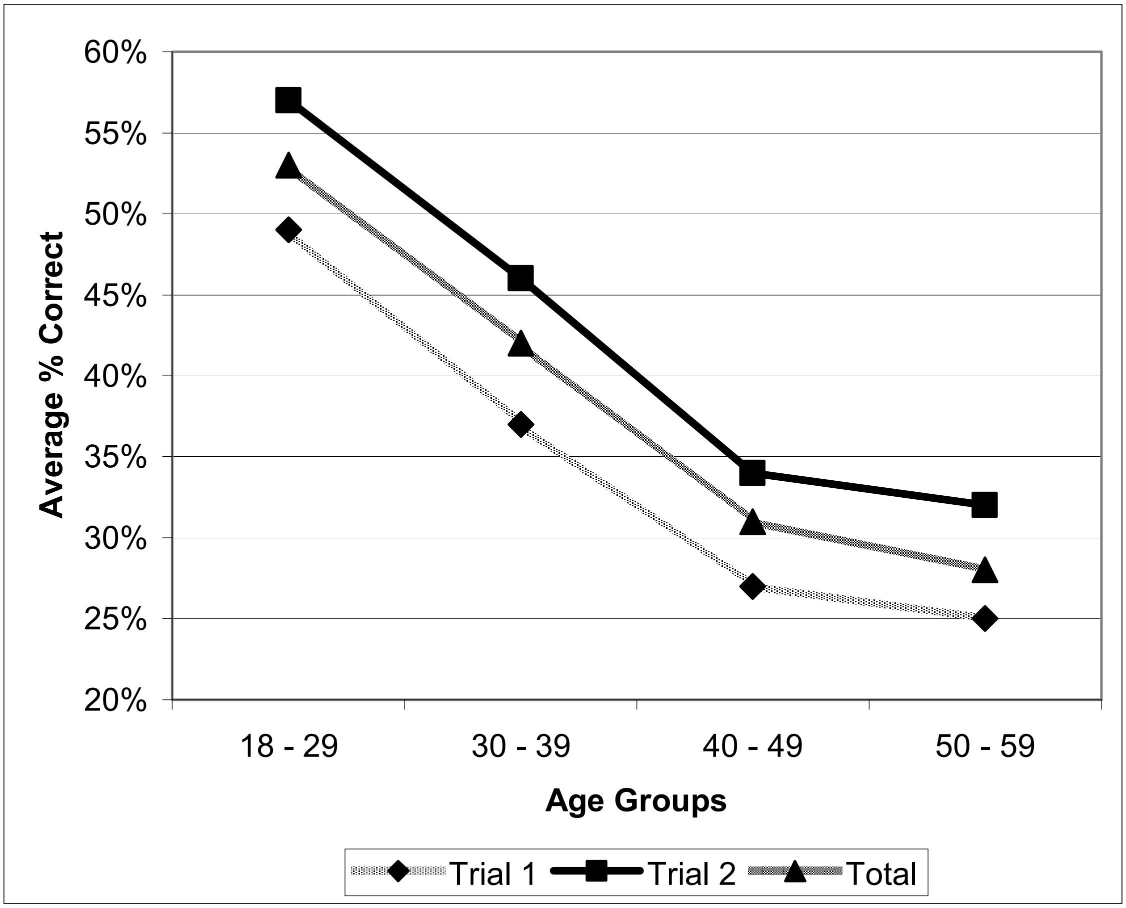

Figure 11. Multitasking Simulation Performance Differences Across Age Groups.

DISCUSSION

Purpose

The purpose of this study was to investigate human performance in an applied work-related multitasking environment. This investigation focused primarily on two classes of variables that could influence performance. First, this study focused on replicating commonly studied task characteristic effects on multitasking performance in lab settings and extending these findings to a practical environment. Second, this study investigated the predictive validity of inferences drawn about performance in a multitasking environment using several individual difference variables.

The first goal of this study centered on understanding how multitasking-environmentally induced task characteristics impacted performance. There were three task characteristic conditions under investigation. (1) Whether or not high demand multitasking environments negatively impact performance relative to monotask and low demand multitasking environments. (2) Whether or not the opportunity to practice performance in a multitasking environment would improve performance. (3) Whether or not an increasingly demanding multitasking environment (in terms of response-stimulus intervals) would lead to a continuous impact on performance.

The second goal of this study was to complete a larger-scale investigation of individual difference predictors of performance in a multitasking environment than has been conducted in past research. This study examined the influence of cognitive ability and eight non-cognitive predictors (extroversion, intellectance, conscientiousness, agreeableness, stress tolerance, polychronicity, and type A behavior pattern: achievement striving and competitiveness). Hypotheses about the strength and direction of each individual difference variables were presented.

Findings

The results described earlier suggest support for most of the hypotheses. All three of the task characteristic hypotheses received support, and five of the nine individual difference hypotheses were supported, with a sixth receiving partial support. In the end, nine out of 12 of the predicted effects received at least partial support (see Table 11 and Table 14).

Hypothesis one predicted that when participants were presented with tasks in a high demand multitasking environment their performance would suffer relative to a low demand multitasking environment. This hypothesis was strongly supported when comparing the lowest demand portion of the multitasking simulation to the highest demand portion (see Figure 8). This finding replicates the "multitasking effect" that has been observed in past laboratory investigations of multitasking (e.g., Rogers & Monsell, 1995). From these data, it appears that the task characteristics present in multitasking environments represent important impediments to performance.

Hypothesis two predicted that practicing performance in a multitasking environment would reduce, but not eliminate the effect of the multitasking task characteristics. This hypothesis was supported. Performance in trial two was much better than performance in trial one (see Figure 8). Further, in comparing specific parallel tasks from the two trials, performance does improve in the second trial for all comparisons. The degree of improvement did vary based on the intensity of the multitasking demand, which has further implications that are discussed below.

Hypothesis three predicted that the interval between the last response and the delivery of the next stimulus would impact performance in a multitasking simulation. Specifically, as this interval reduced, performance would suffer. This hypothesis was supported across most of the intervals in both trial one and trial two. Interestingly, support for this hypothesis varied across multitasking demand (see Figure 9 and Figure 10). In very low demand situations (i.e., large intervals between response and delivery), performance is not impacted by increases. At some point, however, participants "hit a wall" and the interval begins to have a serious impact on performance. This wall was present in both trial one and trial two (that is, both with practice and without).

Hypothesis four was the first individual difference hypothesis. This hypothesis suggested that cognitive ability would be a strong predictor of performance in a multitasking environment. This hypothesis was supported. In fact, the cognitive ability measure was correlated so highly with the multitasking simulation it calls into question whether or not performance in a multitasking environment is a consequence of a separate "multitasking ability" or if "multitasking ability" is better conceptualized as a sub-dimension of 'g'.

The remaining hypotheses suggested that several non-cognitive predictors would either predict or not predict performance in multitasking environment. Four of these hypotheses were supported. Hypothesis five suggested that extroversion would be a weak predictor. Hypotheses seven predicted that conscientiousness would not be associated with performance. Hypothesis eight predicted that agreeableness would not be associated with performance. Hypothesis 12 predicted that polychronicity would have a weak association with performance. All of these predictions were observed as expected; several across multiple performance criteria.

Hypothesis six predicted that intellectance would have a moderate association with performance in a multitasking environment. The magnitude of this relationship was not large enough to be considered moderate, but a significant and important relationship did emerge. The relationship between performance and intellectance was significant, but weak. That said, this predictor was the second strongest predictor of performance behind cognitive ability, which was anticipated. Still, because of the somewhat lower than expected magnitude of the relationship, this hypothesis was only partially supported.

Three hypotheses were not supported. Hypothesis nine predicted that stress tolerance would be a weak predictor of performance in multitasking environment. Hypothesis ten predicted that competitiveness would be a moderate predictor of performance. Hypothesis 11 predicted that achievement striving would have a moderate association with performance. These three predictors were not correlated with performance in a meaningful way. All of the magnitudes were below $r = .10$. All of these predictors also had lower than ideal internal consistency reliability coefficients; these reliability problems may have influenced the observed validity coefficients.

Post-hoc hierarchical regression analyses were performed using the variables that emerged as predictors of performance in a multitasking environment. These analyses suggested that the overwhelming majority of the total variance explained in the performance criteria was accounted for by cognitive ability. The non-cognitive predictors explain a relatively small proportion of the variance in the criteria. These analyses provided further evidence that "multitasking ability" may be better considered as an aspect of cognitive ability, rather than a separate ability.

Further post-hoc analyses investigated differences in performance across gender and age demographics. These findings suggested that only very small differences exist between men and women on performance in a multitasking environment. Age-related differences were more prevalent. Significant performance differences were observed across age groups. As participants get older, performance in a multitasking environment tends to decrease. This finding is consistent with past research (e.g., Salthouse et al., 1998).

Interpretation

Several important trends emerged in these data for both the task characteristic hypotheses and the individual difference hypotheses. The most fundamental trend emerging from the task characteristic hypotheses was that the impact of select task characteristics on performance investigated in past cognitive psychological research is generalizable to work-related multitasking environments. Understanding that these results do hold-up in applied settings has important implications on the directions future research efforts should pursue.

The second trend that emerged out of the task characteristics hypotheses is that there appears to be important and identifiable "levels of impact" as the task characteristics in a multitasking environment vary. These levels of impact emerged largely in the tests of hypotheses two and three. In the tests of hypothesis two, there was an important overall practice effect between trial one and trial two, but the strength of this effect varied when individual task comparisons were examined. The pattern of this variance is interpretable across the trials (see Figure 8). There was a small practice effect on the first tasks of a trial. This is not surprising because the tasks are very easy and there is only a minor multitasking demand. If there is no multitasking demand, then you would not expect an important practice effect given the relative simplicity of the task. That said, as soon as the multitasking demand increases, important practice effects are observed. Under the low- to mid- multitasking demands, practice had a drastic impact on performance. As the trial continued and the multitasking demand became high, practice had a much smaller impact on performance. This trend suggests that practice may help to a point, but when the multitasking demand is too great the effects of practice may disappear.

This trend may explain the mixed results observed in past research. Yeung and Monsell (2003) found that the multitasking effect was robust despite the amount of practice participants engaged in, whereas, Hazeltine et al (2002) and Van Selst (1999) observed beneficial effects of practice on performance in a multitasking environment. The current findings suggest that the impact of practice may depend on the level of intensity of the task characteristics in the multitasking environment.

In sum, there clearly is an important practice effect under low- to medium-intensity multitasking demand, which is consistent with Yeung and Monsell (2003). These data suggest that the impact of practice is somewhat reduced in high-demanding multitasking situations, which is consistent with Hazeltine et al (2002) and Van Selst (1999).

While these results appear to be easily interpretable from the job candidate sample, there is some evidence that practice may have an impact even in demanding multitasking environments when a comparison is made between the candidate sample and the incumbent sample (see Figure 5).

The pattern of results between the candidate sample and incumbent sample are very similar, but there is one important difference relevant to the "practice hypothesis". In trial 2, the job incumbents hit the wall later than the candidate sample. Job candidates begin to have a sharp performance decrease after query three; whereas, job incumbents' performance does not see a sharp decrease until after query four. The job incumbents, on average, will have had much more experience and practice performing in job-related multitasking environments than the job candidates. The incumbent sample consists of participants who successfully completed intense training programs for a job consisting of tasks very similar to the simulation tasks. Further, the incumbent sample participants were on-the-job in a similar multitasking environment for at least three months to be eligible to participate in the study. Therefore, it is safe to presume that the incumbent sample has had considerably more practice in a multitasking environment than the candidate sample.

While more meaningful conclusions could be reached with a future study using a within-subjects design (e.g., retest the job candidate sample after they have been on the job for three months), these data suggest that well-practiced participants may be more resilient to "hitting the wall" in high-demand multitasking environments. Based on this between-subjects comparison of incumbents and candidates, the question of whether or not practice impacts performance even in the most demanding multitasking environments warrants continued research.

In the results for hypothesis three, a trend similar to the trend observed in tests of hypothesis two emerged. In both trials, there was a "wall" hit by participants where the response-stimulus delivery interval led to a drastic increase in performance (see Figure 9 and Figure 10). This finding provides further support that the task characteristics present under various multitasking-demand conditions have different levels of impact. The finding from the tests of this hypothesis show that there is a consistent but small decline in performance under low to medium demand multitasking environments, but when the multitasking demand becomes high the impact on performance drops drastically. Further, while the multitasking demand stays high, performance stays consistently low; it does not improve as the participant remains in the high demand multitasking environment.

Just like the tests of hypothesis two, the "wall" in terms of response-stimulus delivery can also be identified. Interestingly, participants "hit the wall" in terms of practice effects and response-stimulus delivery at practically the same point in the simulation. This consistent trend is important. If organizations can identify whether or not the task characteristics present in the jobs in their organization are intense enough to cause employees to "hit the wall," there could be important implications for both personnel selection and job design in organizations.

The results from the tests of the individual difference hypotheses were relatively easy to interpret. The basic finding from these hypotheses is that cognitive ability appears to be a strong predictor of performance in a multitasking environment, while all other predictors are weak at best.

Several of the weak relationships that emerged were predicted, however, they were relatively close to the predicted threshold of support. For example, support of the extroversion hypothesis is based on a relationship of $r = .12$, and the threshold for a weak relationship is $r = .10$. Confidence intervals based on the sample size of 2711 were constructed to understand how significantly different from the threshold the observed relationships are. This procedure revealed that a relationship of $r = .11$ has a 70% probability of being larger than $r = .10$. A relationship of $r = .12$ has an 85% probability of being larger than $r = .10$. A relationship of $r = .13$ has a 94% probability of being larger than $r = .10$. The probabilities for all other magnitudes are above the conventional 95% standard for statistical significance (these probabilities were calculated by converting z-critical values to p-values). Investigation of these confidence intervals suggests that observed relationships larger than the threshold should be considered as supportive of the predictions with a relatively high degree of confidence.

These results have important implications on how practitioners approach predicting performance in organizations with jobs that have demanding multitasking environments. The relationships from the non-cognitive predictors are important to understand and should not be ignored, but, in terms of practical implications, cognitive ability appears to "rule the day."

Further, the relationships observed for the predictors were consistent with the few studies that have been conducted on the topic. Ishizaka et al. (2001) found that TABP did not predict performance in a multitasking environment. These findings replicated in this study. The relationships observed by Konig et al. (2005) between multitasking performance and extraversion and polychronicity were very similar to the relationships observed in this study. Also, in the Konig et al. (2005) study, cognitive ability emerged as a significantly better predictor of performance in a multitasking environment than the personality variables.

The only interpretation issue with the individual difference hypotheses is the relatively low reliability coefficients observed for stress tolerance, competitiveness, achievement striving, and polychronicity. These lower than ideal reliability results are likely due to the fact that these scales consist of a relatively small numbers of item (e.g., the polychronicity scale is a four-item scale). Regardless of the reason, the low reliability likely impacts the magnitude of the observed correlation coefficients. All of these variables did relate to performance similarly to past research, but because of the reliability issues, more work is needed. For example, polychronicity is a weak predictor of performance in this sample, but would a better measure of the construct yield a stronger relationship? Further, was the lack of relationship observed between stress tolerance and the TABP components due to the fact that there is truly no relationship or is it the case that better predictor measures are needed to identify the nature of the relationship? Answers to these questions need to be addressed in future research.

To gain an understanding of how the low scale reliabilities impacted the observed relationships, a correction for attenuation was performed on the relationships involving the low reliability scales and the key criterion variable, accuracy. The relationship between performance in a multitasking environment and the low reliability scales were improved as follows: (a) stress tolerance went from $r = .04$ to $r = .05$, (b) competitiveness went from $r = .05$ to $r = .06$, (c) achievement striving went from $r = -.03$ to $r = -.04$, and (d) polychronicity went from $r = .13$ to $r = .17$. The results of these corrections do not change support for the hypotheses or the interpretation, but it is important to understand the potential impact low scale reliability coefficients on the observed relationships.

The trends emerging from the age group post-hoc analyses are also interesting. Salthouse et al. (1998) found strong effects on working memory, cognitive ability, and task switching performance based on age group. This study was a lab study and it parsed age into two extreme groupings, so it was not clear that these results would generalize to an applied setting with a more continuous range of ages. The finding that these results did replicate, and the finding that there is such a consistent decline in performance in a multitasking environment, is important. The implications of this finding suggest that practitioners seeking to measure multitasking performance as a predictor of future job performance should proceed with caution.

While the age effect found by Salthouse et al. (1998) replicated in this sample, the results in this study should be interpreted with caution, as there are other plausible drivers of this effect. First, the multitasking simulation is computer-based. It is probably the case that younger job candidates have more computer experience, familiarity, and comfort than older job candidates do. While steps were taken to eliminate computer skill as a confound (see Appendix), these differences could still account for some of the age effect.

Also, there could be ability differences between the age groups due to range restriction differences, rather than actual age-related differences. These candidates were applying for entry-level jobs. It is likely the case that as the age group increases there are fewer high ability candidates in the applicant pool because the high ability candidates already have careers elsewhere. That is, the folks with higher ability in the older age groups already have established careers and tend not to apply to entry-level jobs. From the perspective of the younger age groups, these entry-level jobs are relatively high paying, with lots of room for career development, in a growing and successful company. Consequently, the age differences may be due differences in the host organization's effectiveness in recruiting high ability candidates across age groups.

Implications and Applications

The task characteristic findings illustrate that research conducted on multitasking in controlled laboratory settings is relevant to work-related multitasking environments. While it is likely not the case that all lab-based multitasking findings will replicate across all applied multitasking settings, this study shows that there are some predictable and robust multitasking findings that do replicate. The core characteristic of the multitasking simulation used in this study consist of the need to switch tasks frequently in an environment where the next task switch is uncertain and there are time pressures to complete tasks. These hallmarks of a multitasking situation lead to predictable and interpretable effects that generalize to other situations where these characteristics are present. Therefore, the findings in this study take an important step towards understanding a coherent theory of multitasking.

Delbridge (2000) first identified these characteristics of a multitasking environment. Her theoretical rationale was based on a review of lab-based studies. In her study, she found support for this model. In this study, further support is offered for the usefulness of her model. This study convincingly shows that, even in applied settings, when an environment exists that is typified by frequent task switches, uncertainty, and time pressure; there will be effects on performance similar to those observed in lab-based work.

This point is important as practitioners attempting to measure multitasking through the use of multitasking simulations look for useful research on which to develop their theories. Up until now, applied multitasking work has not tapped into the wealth of laboratory based cognitive multitasking research. This practice has led to a morass of theoretically weak attempts at measuring "multitasking performance." The findings of this study will hopefully serve as an illustration that if practitioners base their multitasking measurement on understanding performance in a multitasking environment with the key characteristics present, then much of the lab-based multitasking research is useful in understanding performance.

Further, the task characteristic findings suggest that performance in a multitasking setting differs from a mono-task setting, but also that performance differs based on the relative intensity of the multitasking setting. This finding may be the most important finding in this study for organizations. The idea that different levels of demand in a multitasking environment have interpretable, measurable, and varying effects on performance can have vast impact on job design and employee selection.

In the tests of hypotheses two and three, participants performed well in low demand multitasking conditions, where time pressures were not intense. As the delivery of tasks increased in frequency, performance was drastically impacted, and when the demand reached a threshold, more interpretable impact on performance was observed. Specifically, these effects suggest that having the opportunity to practice a task in a multitasking environment helps performance, however, there is a point where the multitasking demand becomes so great that practice does not appear to help. Also, the findings show that at the point where the practice effects stop aiding performance is about the same point where the response-stimulus delivery interval has the greatest impact on performance.

From a job design perspective, this collection of findings suggests that it is critical to pay attention to the degree of demand in a multitasking environment. Multitasking is present in many modern day jobs; it is unrealistic to design jobs in our modern organizations that do not have a multitasking component. Organizations should understand, however, that when a multitasking situation becomes too demanding, it is unlikely that training and/or selection interventions will work to improve performance. Rather, it is important that organizations strive to design jobs such that the degree of multitasking demand is manageable, and that the degree of task switching, uncertainty, and time pressure does not exceed their employees' ability to survive. That is, jobs should be designed to make sure that the multitasking demand is not beyond the threshold of the employees.

If organizations have jobs with a multitasking demand that is very high, seeking selection and training strategies that bring in people with higher "multitasking ability" is likely not going to have the desired results. Rather, the organization would be better served to reduce the multitasking demand through job redesign interventions to a level that does not exceed the threshold of the employees' ability. If the threshold is exceeded, performance will likely fluctuate based on the power of the situation and the individual characteristics that predict performance in mono-task and lower demand multitasking environments may not work. However, if jobs can be designed such that the multitasking environment is reasonable, then it is likely that a selection program based on the findings of the individual difference hypotheses could be fruitful in identifying high performers.

If jobs are designed properly, the individual difference findings in this study have important implications for organizations wishing to improve performance in jobs that are typified by a need for multitasking. The results of this study suggest that measures of cognitive ability can explain a considerably large amount of variance in performance in a job-related multitasking simulation. The non-cognitive predictors investigated in this study performed, by and a large, as expected. These results suggest that in the prediction of performance in a multitasking environment, selection programs should center on cognitive ability testing. In fact, the large observed relationship between performance in a multitasking environment and cognitive ability in this study suggests that the two may be similar constructs. In other words, the reading comprehension measure and the multitasking simulation used in this study may simply be measures of different sub-strata in the hierarchy of ‘g’.

There were important mean differences across age groups. This finding suggests that if selection practitioners create simulations that measure performance in a multitasking simulation to be used for employee selection, there may be adverse impact against older job candidates. Lab-based research and the current study suggest that older workers do not perform in multitasking environments as well as younger workers. Selection practitioners should be aware of this situation, and cut scores should be set with caution based on strong evidence of validity.

Future Directions

This study was successful in its goal to understand more about the nature of performance in applied multitasking environments, both in terms of the generalizability of lab-based findings and the nature of individual difference predictor relationships, but there is still much research to be done. The findings from this study identify several avenues for future research on the topic of applied multitasking performance.

First, future research should continue to investigate additional findings from lab-based cognitive psychological research on multitasking. For example, this study focuses on the task switching aspects of multitasking performance. While there is some dual processing required in the multitasking simulation, the main manipulations center on switching tasks. Future research should seek to understand the levels of dual processing that occur in modern day multitasking jobs. If these levels can be identified and measured, it would be useful to attempt to replicate the findings of dual task research to applied multitasking settings.

Future research should also continue to investigate performance differences across varying degrees of multitasking demand. The current study does seem to identify some important differences at varying levels of demand in the multitasking simulation, but there is still much to be investigated. Research should examine the aspects and levels of the situation that causes participants to "hit the wall". This line of research would have important implications for job design and employee selection.

Researchers should look at differences in prediction across different levels of multitasking demand. For example, researchers should investigate whether or not cognitive ability predicts performance differentially at different levels. In the current data set, there is a slight difference. The correlation between medium demand queries and cognitive ability is .64, and the correlation between high demand queries and cognitive ability is .58. This difference is too small to be meaningful, but this simulation was not ideally set-up to measure these prediction differences. Researchers interested in investigating this research question should use simulations with more measurement for each respective level of multitasking demand with a wider range of stimulus delivery frequency.

Similarly, researchers should investigate the influence of individual differences across different levels of multitasking demand. For example, researchers should investigate whether or not people with varying levels of cognitive ability "hit the wall" differently. To fully answer this research question the multitasking tool should include more measurement across task characteristic demands and sampling strategies ensuring that different "levels" of relevant individual differences are represented. That said, the data from this study provide preliminary evidence that there may be some differences. For example, Figure 12 shows the query means for participants outside of one standard deviation from the mean on the cognitive ability test on the high and low ends of the distribution. This chart suggests that both groups do "hit the wall", but the effect is somewhat muted for the low ability group. Rather, the performance decline in the low ability group is more gradual and consistent in the most intense portions of the multitasking simulation. Regardless of ability, there does appear to be a significant impact on performance at the most intense levels of multitasking demand.

These data also have implications for the "practice hypothesis." Figure 12 shows that the difference between trial 1 and trial 2 is much larger for the high ability group than the low ability group when the multitasking demand is intense. This finding suggests that while redesigning the workspace to a more manageable level will help, hiring high ability job candidates will still have benefits. High ability people may be better suited to learn how to perform in environments with intense multitasking demands and their absolute level of performance will be higher despite the multitasking demand. Further, these data suggest that not only will low ability people be sensitive to changes in the multitasking environment, but also they will not be able to improve with practice as well as high ability people.

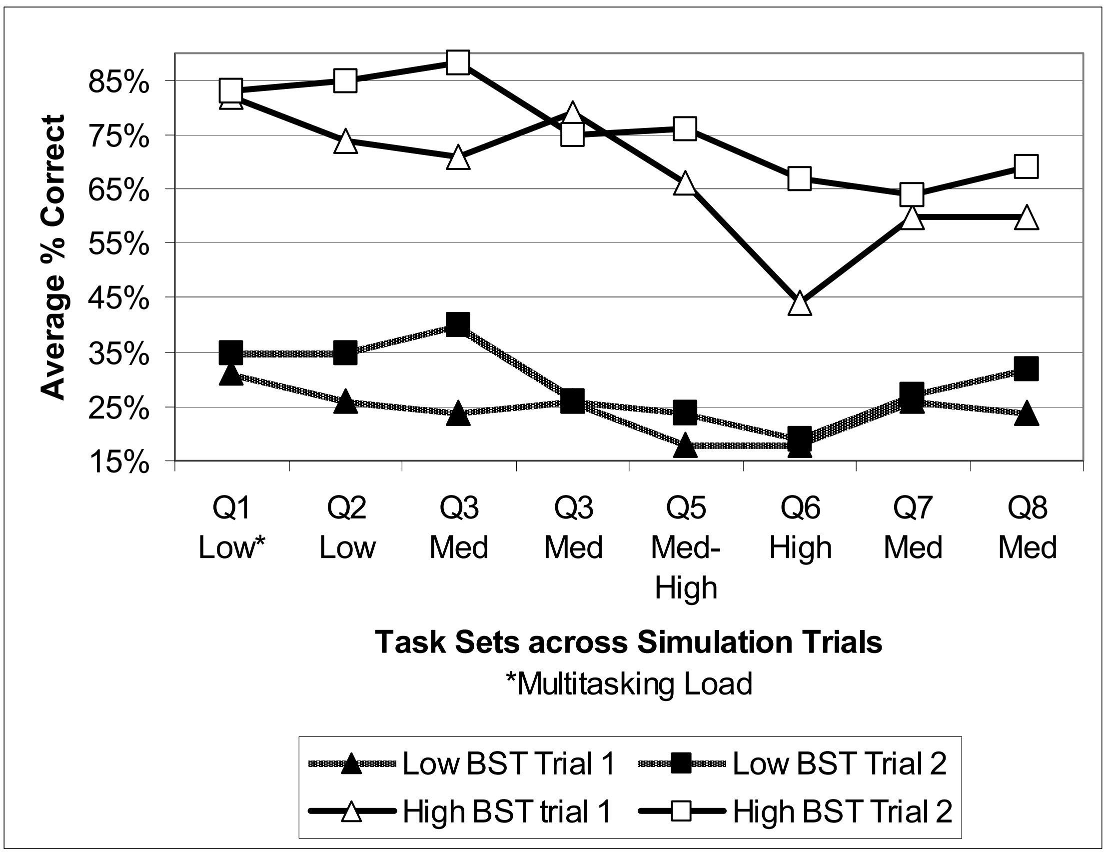

Figure 12. Patterns of Performance Across Trial One and Trial Two in High- and Low-Cognitive Ability Groups.

This study also highlights the need for continued research on individual difference predictors of performance in a multitasking environment. While this study provides strong support for the influence of cognitive ability on performance in a multitasking environment, there are still unanswered questions about the non-cognitive predictors. Particularly, the nature of the relationship between performance and the scales with lower reliability warrant further research. Future research should investigate whether or not the lack of relationship between TABP and stress tolerance with performance in a multitasking environment holds up with better measures of these constructs. The low reliability of these scales makes it difficult to trust the results observed in this study.

Further, continued work on the nature of the relationship between performance in a multitasking environment and polychronicity is warranted. The polychronicity scale used in this study correlated highly with another measures of polychronicity in a different study, but this scale consistently has lower than ideal reliability. The magnitudes of the relationships observed between the criteria and polychronicity were in line with past research, but it is possible that the true strength of the association has yet to be discovered. In past research on this topic, there have been problems with the criterion measure and with small sample sizes. In this study, the criteria are solid and the sample is large, but the predictor measure is not adequate. Even with these problems across studies, the relationship is robust enough to be present consistently. Perhaps future research where these problems are not present will fill in the remaining gap of understanding about the magnitude of the polychronicity-multitasking performance association.

While the findings in this study suggest that the non-cognitive individual difference predictors do not explain much performance variance in a multitasking environment, this effect could be due to the nature of the task. It is possible that in other multitasking environments some of the non-cognitive predictors would explain more performance variance. Performance in the multitasking simulation in this study is dependent upon learning novel tasks in a relatively short period of time and answering questions based on the content presented. Cognitive ability could influence performance in this particular setting because of the cognitive demands of the task rather than the multitasking demands of the situation.

A worthwhile avenue of research would be to extend the findings of this study to multitasking environments with less cognitive demand to understand whether or not the non-cognitive predictors explain more variance on these tasks. Lieberman and Rosenthal (2001) discuss the role of extraversion in interpersonal multitasking situations. These researchers explained that extraverts are better suited to attend to and interpret simultaneously occurring verbal and non-verbal cues in interpersonal settings. Future research should investigate the role of extraversion in multitasking environments with increased emphasis on interpersonal tasks.

Another worthwhile extension of these findings would be to investigate performance in a multitasking environment where the are chosen by the participant. In the current multitasking simulation, the “next” task is fixed, or defined largely by the situation, not the participant. This lack of choice could impact how some of the non-cognitive predictors influence performance. For example, Persing (1999) discussed the importance of personal agency on the effect of polychronicity. This author pointed out that the power of the situation can trump individual differences in polychronicity. Persing (1999) suggested that polychrons will not exhibit a preference for polychronic time if the need for multitasking performance is driven by the situation. Future research should investigate whether or not polychronicity exerts more influence on performance in multitasking situations involving more personal agency.

Also, future research should investigate whether or not other cognitive abilities differentially predict performance in a multitasking environment. It is likely that any construct valid measure of ‘g’ should work, but this study only investigated reading comprehension. Of particular interest are information processing measures, which tend to load on ‘g’ (e.g., short-term memory). Typically, studies investigating incremental validity of psychometric ‘g’ and information processing turn up only small differences. However, these studies have never investigated incremental validity in a multitasking environment. It is possible that information-processing measures explain performance variance above and beyond psychometric ‘g’ measures due to the increased demand on the information processing system in multitasking environments. This could be an important new avenue for applied cognitive ability research.

Similarly, research should investigate whether creating simulations like the tool used in this study to be used in selection is worthwhile given the large relationship with psychometric ‘g’. Would a simulation-based measure of performance in a multitasking environment explain more variance than traditional cognitive ability testing? Such a study should be careful to choose criteria that reflect the multitasking nature of the job in question, as well as other typical criteria. Prediction differences could have important implications on the landscape of employment testing.

Further, are there differences in adverse impact for tools such as the multitasking simulation relative to traditional cognitive ability testing? The current results suggest this type of simulation may not have an adverse impact against females, but probably does have a disparate impact against older job candidates. These potential group differences are important to understand.

One advantage of these types of multitasking simulations is that they can be constructed to be extremely face valid. Further, they can provide a realistic job preview to job candidates about the demands of the job. Given these considerations, it could be the case that even if multitasking simulations do not explain performance variance above and beyond cognitive testing for traditional criterion variables, perhaps simulations such as these would improve prediction of criteria such as turnover.

The bounds of future research in the area of applied multitasking performance are constrained only by the imagination of the researcher. At this point, there has not been a lot of work done. This dissertation is a vivid illustration that this area is a worthy and important area of research. The questions answered in this study only scratch the surface of what there is to learn. Hopefully, researchers will continue to develop and test theories to gain a greater understanding of the dynamics of performance in work-related multitasking environments.

REFERENCES

Anderson, T. N., & Kida, T. E. (1985). The effect of environmental uncertainty on the association of expectancy attitudes, effort, and performance. The Journal of Social Psychology, 125, 631 – 636.

Barrick, M, & Mount, M. (1991). The big five personality dimensions and job performance: A meta-analysis. Personnel Psychology, 44, 1 – 26.

Barrick, M., Mount, M, & Strauss, J. (1994). Antecedents of involuntary turnover due to a reduction of force. Personnel Psychology, 47, 515 – 535.

Bazana, P. G., & Stelmack, R. M. (2002). Intelligence and information processing during an auditory discrimination task with backward masking: An event-related potential analysis. Journal of Personality and Social Psychology, 83 (4), 998 – 1008.

Benabou, C. (1999). Polychronicity and temporal dimensions of work in learning organizations. Journal of Managerial Psychology, 14, 257 – 268.

Ben-Shakhar, G., & Sheffer, L. (2001). The relationship between the ability to divide attention and standard measures of general cognitive abilities. Intelligence, 29, 293 – 306.

Bemudez, J., Perez-Garcia, A. M., & Sanchez-Elvira, A. M. (1990). Type A behavior pattern and attentional performance. Personality and Individual Differences, 11, 13 – 18.

Bluedorn, A. C., & Denhardt. (1988). Time and organizations. Journal of Management, 14(2), 299 – 320.

Bluedorn, A. C., Kalliath, T. J., Strube, M. J., & Martin, G. D. (1999) Polychronicity: A fundamental dimension of organizational culture. Journal of Managerial Psychology, 3, 205 – 230.

Bluen, S. D., Barling, J., & Burns, W. (1990). Predicting sales performance, job satisfaction, and depression using the achievement strivings and impatience-irritability dimensions of Type A behavior. Journal of Applied Psychology, 75, 212 – 216.

Bond, M. H., Nakazato, H. S., & Shiraishi, D. (1975). Universality and distinctiveness in dimensions of Japanese person perception. Journal of Cross-Cultural Psychology, 6, 346 – 355.

Borkenau, P., & Ostendorf, F. (1991). A questionnaire for assessing five robust personality factors. Diagnostica, 37, 29 – 41.

Borkenau, P., & Ostendorf, F. (1993). The NEO-Five Factor Inventory (NEO – FFI) after Costa & McCrae. Gottingen, Germany: Hogrefe.

Bratfisch, O., & Hagman, E. (2003). Simultankapazitat/Multi-tasking (SIMKAP) Version 24.00: Handanweisung [Simultaneous Capacity/multi-tasking (SIMKAP) release 24.00: Manual]. Modling, Austria: Schuhfried.

Bratfisch, O., Hagman, E., & Puhr, U. (2002). Simultaneous Capacity (Release 23.00): Manual. Modling, Austria: Schuhfried.

Bremner, S. (2005). The perils of multitasking: Is less really more? Health Sciences Information Service, No. 47, Novemeber, online – www.lib.uaaalaska.edu/hsis/

Broadbent, D. E. (1958). Perception and Communication. New York: Pergamon Press.

Burnam, M. A., Pennebaker, J. W., & Glass, D. C. (1975). Time consciousness, achievement striving, and the Type A coronary-prone behavior pattern. Journal of Abnormal Psychology, 84, 76 – 79.

Carroll, J. B. (1993). Human Cognitive Abilities: A Survey of Factor-Analytic Studies. Cambridge, England: Cambridge University Press.

Carver, C. S., Coleman, A. E., & Glass, D. C. (1976). The coronary-prone behavior pattern and the suppression of fatigue on a treadmill test. Journal of Personality and Social Psychology, 33, 460 – 466.

Carver, C. S., & Glass, D. C. (1978). Coronary-prone behavior pattern and interpersonal aggression. Journal of Personality and Social Psychology, 36, 361 – 366.

Carver, C. S., Peterson, L. M., Follansbee, D. J., Scheier, M. F. (1983). Effects of self-directed attention on performance and persistence among persons high and low in test anxiety. Cognitive Therapy and Research, 7, 333 – 354.

Cohen, J. (1988). Statistical power analysis for the behavioral sciences (2nd ed.). Hillsdale, NJ: Erlbaum.

Conte, J. M., & Gintoft, J. N. (2005). Polychronicity, big five personality dimensions, and sales performance. Human Performance, 18(4), 427 – 444.

Conte, J. M., & Jacobs, R. R. (2003). Validity evidence linking polychronicity and big five personality dimensions to absence, lateness, and supervisory performance ratings. Human Performance, 16(2), 107 – 129.

Conte, J. M., Rizzuto, T. E, & Steiner, D. D. (1999). A construct-oriented analysis of individual-level polychronicity. Journal of Managerial Psychology, 14, 269 – 287.

Cornelius, S. W., Willis, S. L., Nesselroade, J. R., & Baltes, P. B. (1983). Convergence between attention variables and factors of psychometric intelligence in older adults. Intelligence, 7, 253 – 269.

Costa, P. T. Jr., & McRae, R. R. (1988). From catalog to classification: Murray's needs and the five-factor model. Journal of Personality and Social Psychology, 55, 258 – 265.

Cotte, J., Ratneshwar, S. (1999). Juggling and hopping: what does it mean to work polychronically? Journal of Managerial Psychology, 14, 184 – 204.

Damos, D., & Smist, T. (1982). Individual differences in multi-task response strategies. Aviation, Space, and Environmental Medicine, 53, 1177 – 1181.

Deary, I. J., & Stough, C. (1996). Intelligence and inspection time: Achievements, prospects, and problems. American Psychologist, 51, 599 608.

De Jong, R. (1995). The role of preparation in overlapping-task performance. The Quarterly Journal of Experimental Psychology.

De la Casa, L. G., Gordillo, J. L., Mijias, L. J., Rangel, F., & Romero, M. F. (1998). Attentional strategies in Type A individuals. Personality and Individual Differences, 24, 59 – 69.

Delbridge, K. A. (2000). Individual differences in multi-tasking ability: Exploring a nomological network. Unpublished Dissertation; Michigan State University.

Dickman, S. (1985). Impulsivity and perception: Invidual differences in the processing of the local and global dimensions of stimuli. Journal of Personality and Social Psychology, 48(1), 133 – 149.

Dickman, S., & Meyer, D. E. (1988). Impulsivity and speed-accuracy tradeoffs in information processing. Journal of Personality and Social Psychology, 54 (2), 274 – 290.

Digman, J. M. (1990). Personality Structure: Emergence of the five factor model. Annual Review of Psychology, 41, 417 – 440.

Edwards, J. R., Baglioni, A. J., Jr., & Cooper, C. L. (1990). Examining the relationships among self-report measures of the Type A behavior pattern: The effects of dimensionality, measurement error, and differences in underlying constructs. Journal of Applied Psychology, 75, 440 – 454.

Fiske, D. W. (1949). Consistency of factorial structures of personality ratings from different sources. Journal of Abnormal Social Psychology, 44, 329 – 344.

Friedman, M., & Rosenman, R. H. (1959). Association of a specific overt behavior pattern with blood and cardiovascular findings. Journal of the American Medical Association, 169, 1286 – 1296.

Friedman, M., & Rosenman, R. H. (1974). Type A Behavior and Your Heart. New York, NY: Knopf.

Gardner, D. G. (1986). Activation theory and task design: An empirical test of several new predictions. Journal of Applied Psychology, 71, 411 – 418.

Gardner, D. G., & Cummings, L. L. (1988). Activation theory and task design: Review and reconceptualization. In B. M. Staw and L. L. Cummings (Eds.), Research in Organizational Behavior, 10, Greenwich, CT: JAI Press Inc.

Goldberg, L. R. (1981). Language and individual differences: The search for universals in personality lexicons. In Wheeler, L. (Ed.), Review of Personality and Social Psychology, (Vol. 2), pp. 141 – 166. Beverly Hills, CA: Sage Publications.

Goldstein, H. W., Zedeck, S. Goldstein, I. L. (2002). g: Is this your final answer? Human Performance, 15.5, 123 – 142.

Gopher, D. (1982). A selective attention test as a predictor of success in flight training. Human Factors, 24, 173 – 183.

Gopher, D., Armony, L., & Greenspan, Y. (2000). Switching tasks and attention policies. Journal of Experimental Psychology: General, 129, 229 – 308.

Gopher, D. & Kahneman, D. (1971). Individual differences in attention and the prediction of flight criteria. Perceptual and Motor Skills, 33, 1335 – 1342.

Gottfredson, L. S. (1997). Why g matters: The complexity of everyday life, Intelligence, 24, 79 – 132.

Gottfredson, L. S. (2002). Where and why g matters: Not a mystery. Human Performance, 15.5, 25 – 46.

Gray, J. A. (1987). The Psychology of Fear and Stress, (Second Edition). Cambridge, England: Cambridge University Press.

Gross, B. (1987). Time scarcity: Interdisciplinary perspectives and implications for consumer behaviors. In J. N. Sheth and E. C. Hirschman (Eds), Research in Consumer Behavior, Vol 2, JAI Press, Greenwich, CT, pp 1 – 54.

Guzy, L. T., & Axelrod, S. (1972). Interaural attention shifting as response. Journal of Experimental Psychology, 95, 290 – 294.

Hall, E. T. (1959). The Silent Language. New York, NY: Anchor Press.

Hall, E. T. (1983). The Dance of Life: The Other Dimension of Time. Garden City, NY: Anchor Press.

Hazeltine, E., Ruthruff, E. & Remington. (2006). The role of input and output modality pairings in dual-task performance: Evidence for context-dependent central interference. Cognitive Psychology, 52, 291 – 345.

Hazeltine, E., Teague, D., & Ivry, R. B. (2002). Simultaneous dual-task performance reveals parallel response selection after practice. Journal of Experimental Psychology, 28, 527 – 545.

Hein, G., & Schubert, T. (2004). Aging and input processing in dual-task situation. Psychology and Aging, 19, 416 – 432.

Helmreich, R. L., Spence, J. T., & Pred, R. S. (1988). Making it without losing it: Type A, achievement motivation, and scientific attainment revisited. Personality and Social Psychology Bulletin, 14, 495 – 504.

Hunt, E. (1978). Mechanics of verbal ability. Psychological Review, 85, 109 – 130.

Hunter, J. E., & Hunter, R. F. (1984). Validity and utility of alternative predictors of job performance. Psychological Bulletin, 96, 72 – 98.

Ishizaka, K., Marshall, S. P., & Conte, J. M. (2001). Individual differences in attentional strategies in multitasking situations. Human Performance, 14(4), 339 – 358.

Jerslid, A. T. (1927). Mental set and shift. Archives of Psychology. (Whole number 89).

Joslyn, S. L. & Hunt, E. (1998). Evaluating individual differences in response to emergency situations. Journal of Experimental Psychology: Applied, 4, 16 – 43.

Kahneman, D., Ben-Ishai, R., & Lotan, M. (1973). Relation of a test of attention to road accidents. Journal of Applied Psychology, 58, 113 – 115.

Kaufman-Scarborough, C., & Lindquist, J. D. (1999). Time management and polychronicity: Comparisons, contrasts, and insights for the workplace. Journal of Managerial Psychology, 14, 288 – 312.

Kinney, T. B. (2004). Unpublished Benchmarking Study on the state of the practice in multitasking measurement.

Kinney, T. B., Phillips, S., Michel, R., Higgins, M. (2005). Unpublished Job Analysis for Sales Agents, Service Counselors, Claims Service Representatives, and Telephone Claims Representatives/Direct Handlers for a leading automobile insurance provider.

Klapp, S. T., Boches, C. A., Trabert, M. L., & Logan, G. D. (1991). Automatizing alphabet arithmatic: II. Are there practice effects after automaticity is achieved? Journal of Experimental Psychology: Learning, Memory, and Cognition, 17, 196 – 209.

Konig, C. J., Buhner, M., & Murling, G. (2006). Working memory, fluid intelligence, and attention are predictors of multitasking performance, but polychronicity and extraversion are not. Human Performance, 18(3), 243-266.

Kray, J., Li, K. Z. H., & Lindenberger, U. (2002). Age-related changes in task switching components: The role of task uncertainty. Brain and Cognition, 49, 363 – 381.

Kyllonen, P. C., & Christal, R. E. (1990). Reasoning ability is (little more than) working memory capacity? Intelligence, 14, 389 – 433.

Laabs, G. J., & Stager, P. (1976). Monitoring the information-processing demands of attention shifting. Canadian Journal of Psychology, 30, 47 – 54.

Landy, F. J., Rastegary, H., Thayer, J., & Colvin, C. (1991). Time urgency: The construct and its measurement. Journal of Applied Psychology, 76, 644 – 657.

Lieberman, M. D., & Rosenthal, R. (2001). Why introverts can't always tell who likes them: Multitasking and nonverbal decoding. Journal of Personality and Social Psychology, 80(2), 294 – 310.

Lien, M., Procter, R. W., & Ruthruff, E. (2003). Still no evidence for perfect timesharing with two ideomotor-compatible tasks: A reply to Greenwald (2003). Journal of Experimental Psychology, 29, 1267 – 1272.

Lindenberger, U., Marsiske, M., Baltes, P. B., (2000). Memorizing while walking: Increase in dual-task costs from young adulthood to old age. Psychology and Aging, 15, 417 – 436.

Logan, G. D. (1988). Towards and instance theory of automaticity. Psychological Review, 95, 492 – 527.

Logan, G. D., & Klapp, S. T. (1991). Automatizing alphabet arithmatic: I. Is extended practice necessary to produce automaticity? Journal of Experimental Psychology: Learning, Memory, and Cognition, 17, 179 – 195.

Mann, L. (1992). Stress, Affect, and Risk Taking. Chichester: John Wiley & Sons.

Matthews, K. A., & Brunson, B. I. (1979). Allocation of attention and type A coronary prone behavior. Journal of Personality and Social Psychology, 37, 2081 – 2090.

Matthews, G., Davies, D. R., & Lees, J. L. (1990). Arousal, extraversion, and individual differences in resource availability. Journal of Personality and Social Psychology, 59 (1), 150 – 168.

Matthews, G., Jones, D. M., & Chamberlain, A. G. (1989). Interactive effects of extraversion and arousal on attentional task performance: Multitple resources or encoding processes? Journal of Personality and Social Psychology, 56 (4), 629 – 639.

McRae, R. R. (1996). Social consequences of experiential openness. Psychological Bulletin, 120, 323 – 337.

Meiran, N. (1996). The reconfiguration of processing made prior to task performance. Journal of Experimental Psychology: Learning, Memory, and Cognition, 22, 1423 – 1442.

Meyer, D. E., & Kieras, D. E. (1997). A computational theory of executive cognitive processes and multiple-task performance: I. Basic Mechanisms. Psychological Review, 104, 3 – 65.

Mintzberg, H. M. (1973). The nature of managerial work. New York: Harper & Row, Publishers.

Moch, A. (1984). Type A and Type B behavior patterns, task type, and sensitivity to noise. Psychological Medicine, 14, 643 – 646.

Mount, M., & Barrick, M. (1995). The big five personality dimensions: Implications for research and practice in human resources management. Research in Personnel and Human Resources Management, 13, 153 – 200.

Moustafa, K. S., Bhagat, R. S., Babakus, E. (2005). A cross-cultural investigation of polychronicity: A study of organizations in three countries. Presented at the Midwest Academy of Management Annual Conference, April 2005, Chicago, IL.

Murphy, K. R. (1989). Is the relationship between cognitive ability and job performance stable over time? Human Performance, 2, 183 – 200.

Murphy, K. R. (1996). Individual differences and behavior in organizations: Much more than g. In K. R. Murphy (Ed.), Individual Differences and Behavior in Organizations (pp. 3 – 30). San Francisco: Jossey-Bass.

Murphy, K. R. (2002). Can conflicting perspectives n the role of g in personnel selection be resolved? Human Performance, 15.5, 173 – 186.

Nettelbeck, T., & Lally, M. (1976). Inspection time and measured intelligence. British Journal of Psychology, 67, 17 – 22.

Norman, W. T. (1963). Toward an adequate taxonomy of personality attributes: Replicated factor structure in peer nomination personality ratings. Journal of Abnormal and Social Psychology, 66, 574 – 583.

Nygren, T. E. (1997). Framing of task performance strategies: Effects on performance in a multiattribute dynamic decision making environment. Human Factors, 1997, 425 – 437.

O'Hare, D. (1997). Cognitive ability determinants of elite pilot performance. Human Factors, 39, 540 – 552.

Persing, D. L. (1999). Managing in polychronic times: Exploring individual creativity and performance in intellectually intensive venues. Journal of Managerial Psychology, 14, 358 – 373.

Recarte, M. A., & Nunes, L. M. (2003). Mental workload while driving: Effects on visual search, discrimination, and decision making. Journal of Experimental Psychology: Applied, 9, 119 – 137.

Ree, M. J., & Carretta, T. R. (2002). g2K. Human Performance, 15.5, 3 – 24.

Reeve, C. L., & Hakel, M. D. (2002). Asking the right questions about g. Human Performance, 15.5, 47 – 74.

Roberts, R. D., Beh, H. C., & Stankov, L. (1988). Hick's law, competing-task performance, and intelligence. Intelligence, 12, 111 – 130.

Rogers, R. D., & Monsell, S. (1995). Costs of a predictable switch between simple cognitive tasks. Journal of Experimental Psychology: General, 124, 207 – 231.

Russell, D., & Sacco, J. (2006). Selecting workers who can multitask: The state of our practice. Presentation at the 21st annual meeting for the Society of Industrial/Organizational Psychologists. Dallas TX.

Salthouse, T. A., Fristoe, N., McGuthry, K. E., & Hambrick, D. Z. (1998). Relation of task switching to speed, age, and fluid intelligence. Psychology and Aging, 13, 445 – 461.

Santosus, M. (2003). Why more is less: Reality Bytes. CIO Magazine. September, 15, online – www.cio.com/archive/091503/reality.html.

Saucier, G., & Ostendorf, F. (1999). Hierarchical subcomponents of the big five personality factors: A cross-language replication. Journal of Applied Psychology, 76(4), 613 – 627.

Shaurbroeck, J. & Williams, S. (1993). Type A behavior and escalating commitment. Journal of Applied Psychology, 78, 862 – 867.

Schmidt-Atzert, L., & Deter, B. (1993). Intelligence and training success: A study of predictive validity of the German Intelligence Test I-S-T 70. Zeitschrift fur arbetis- und Organisationpsychologie, 37, 52 – 63.

Schmidt, F. L. (2002). The role of general cognitive ability and job performance: Why there cannot be a debate. Human Performance, 15.5, 187 – 210.

Schmidt, F. L., & Hunter, J. E. (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. Psychological Bulletin, 124, 262 – 274.

Schmitt, N., Gooding, R. Z., Noe, R. A., & Kirsch, M. (1984). Meta analyses of validity studies published between 1964 and 1982 and the investigation of study characteristics. Personnel Psychology, 37, 402 – 422.

Sohn, M., & Anderson, J. R. (2001). Task preparation and task repetition: Two-component model of task switching, 130, 764 – 778.

Spearman, C. (1904). "General Intelligence" objectively defined and measured. American Journal of Psychology, 15, 201 –293.

Spence, J. T., Helmreich, R. L., & Pred, R. S. (1987). Impatience versus achievement striving in the Type A pattern: Differential effects on students' health and academic achievement. Journal of Applied Psychology, 72, 522 – 528.

Stankov, L. (1988). Single tests, competing tasks and their relationship to the broad factors o intelligence. Personality and Individual Differences, 9, 25 – 33.

Stankov, L. (1989). Attentional resources and intelligence: A disappearing link. Personality and Individual Differences, 9, 25 – 33.

Stauffer, J. M., Ree, M. J., & Carretta, T. R. (1996). Cognitive-components tests are not much more than g: An extension of Kyllonen's analyses. Journal of General Psychology, 123, 193 – 205.

Strayer, D. L., Drews, F. A., & Johnston, W. A. (2003). Cell phone-induced failures of visual attention during simulated driving. Journal of Experimental Psychology: Applied, 9, 23 – 32.

Strayer, D. L., & Johnston, W. A. (2001). Driven to distraction: Dual-task studies of simulated driving and conversing on a cellular phone. Psychological Science, 462 – 466.

Tenopyr, M. L. (2002). Theory versus reality: Evaluation of g in the workplace. Human Performance, 15.5, 98 – 108.

Tombu, M., & Jolicoeur, P. (2004). Virtually no evidence for virtually perfect time-sharing. Journal of Experimental Psychology, 30, 795 – 810.

Tupes, E. C., & Christal, R. E. (1961). Recurrent Personality Factors Based on Ratings. (ASD-TR-61-97). Lackland Air Force Base, TX: Aeronautical Systems Division, Personnel Laboratory.

Van Selst, M., Ruthruff, E., & Johnston, J. C. (1999). Can practice effects eliminate the psychological refractory period effect. Journal of Experimental Psychology: Human Perception and Performance, 25, 1268 – 1283.

Weber, R. J., Blagowsky, J., & Mankin, R. (1982). Switching time between overt and covert speech: Generative attention. Memory and Cognition, 10, 546 – 553.

Weber, R. J., Burt, D. B., & Noll, N. C., (1986). Attention switching between perception and memory. Memory and Cognition, 14, 238 – 245.

Wood, R. E. (1986). Task complexity: Definition of the construct. Organizational Behavior and Human Decision Processes, 37, 60 – 82.

Yeung, N., & Monsell, S. (2003). The effects of recent practice on task switching. Journal of Experimental Psychology: Human Perception and Performance, 29(5), 919 – 936.

Zaldivar, K. D. (2006). Multi-tasking: An overview of the research. Presentation at the 21st annual meeting for the Society of Industrial/Organizational Psychologists. Dallas TX.

Zimmerman, P., & Fimm, B. (1993). Test Battery for Attentional Performance (TAP) version 1.5: Manual. Herzogenrath, Germany: Psytest.

APPENDIX

<u>The Development of a Realistic, High-Fidelity, Multitasking Performance Simulation</u>

In early 2004, I was put in charge of an initiative to explore the use of a multitasking job simulation as part of the entry-level selection system of an industry leading insurance provider. Based on several informal conversations with managers in the organization, a lack of multitasking "ability" was perceived to be leading to failure and a higher than ideal turnover rate. Initially, the goal of this initiative was to find a vendor developed measure of multitasking to apply as part of the hiring process.

To select a tool, the SIOP website "Consultant Locator" feature was used to identify potential consultants that cater to this type of selection. Approximately, 50 consulting firms were identified and a brief survey was sent to each vendor to identify the relevance of their tools to the host organization's needs. As a result of this benchmarking study, several trends emerged among the state of the practice in multitasking measurement: (a) there are only a handful of firms that offer multitasking simulations, and (b) most of these simulations were not good enough to consider purchasing (Kinney, 2004). This benchmarking study found that most multitasking tools can be classified as one of two types: data entry measures and indirect measures (Kinney, 2004).

The data entry measures involve simple tasks where information is presented audibly that must be transcribed into a computer system. These tools tend to be highly contaminated by typing skills and typically involve very little task switching and uncertainty. The indirect measures of multitasking are slightly better in that they frequently do involve true multitasking, but this construct is not actually measured. That is, several consultants offer high fidelity job simulations where many skills and abilities are measured and in order to do well on all dimensions, the candidate must be able to multitask well too.

Because neither of these options was deemed attractive, the results of the benchmarking study convinced the host organization to pursue in-house development of its own multitasking simulation based on the existing research literature on the topic. The goal was to create a theory-driven selection simulation, where performance is measured in a fast paced, demanding, computer-based environment. Unlike the tools uncovered in the benchmarking study (Kinney, 2004), this tool had to present realistically complex tasks that adhere to the characteristics of a multitasking environment as defined by the research literature. The tool had to (a) return a score of performance in a multitasking environment that is consistent with how it has been measured in the research literature, and (b) minimize confounds with typing skill and computer skill as much as possible.

The development of the tool started at the beginning of 2005. A literature review uncovered the characteristics of a multitasking environment discussed in the introduction (e.g., frequent task switching, task switch uncertainty, and demanding time pressures). The next step was to investigate the jobs in the host organization to gain an understanding of how these environmental characteristics are manifested in the jobs of the host organization. To gain this understanding several job observations, focus groups, and other job analytic techniques were employed (Kinney et al 2005). After reaching an understanding of the demands on the job and how the key features of a multitasking environment created these demands, the measurement approach logically followed.

I developed a background story for the simulation designed to be face valid for the candidates. In this simulation the candidate must manage customer information for a fictitious supplemental medical insurance company. The simulated customers must be directed to the sales department, the service department, or the claims department. The candidate must identify key information from the phone call or email and enter it into the system. All of the ideas, content, manipulations, and stimulus materials in the simulation, other than the actual Flash programming, were created by the author.

In the simulation the candidate has to switch tasks continually. S/he must choose emails and phone calls (both are referred to as "queries") to process, perform various mousing, typing, and computer operations, read emails, and attend to audio information. The task switching demand becomes heavier as the simulation progresses.

The pace that calls and emails are delivered to the candidate creates the need for frequent task switching and the frequency increases as the simulation progresses from a relatively low task switch demand in the beginning to a high task switch demand towards the end. The first six calls/emails in a trial are delivered in decreasing intervals starting at 55 seconds apart than delivered 5 seconds sooner until the sixth call/email, which is delivered 35 seconds after the previous call/email. The final two calls/emails are delivered 40 seconds apart. Because of the demanding nature of the tasks, the candidates do not catch on to the pattern. Candidates do realize that the pace is increasing, but too much of their resources are allocated on completing the task to uncover any pattern of task delivery. In fact, most participants appear to be so overloaded by the tasks they are working on that it takes time to even register when the next call/email arrives.

Time pressure is created by two on-screen timers. When a call or email arrives in the candidate's "query bank", s/he has one-minute to open it and once opened, s/he has one minute to process it. As the trial progresses and queries stack in the query bank, the workload becomes increasingly difficult making the time pressures increasingly salient.

Finally, attention was paid to the modality of stimulus presentation and response medium. Researchers have found that the cost of dual task performance tends to depend on how the stimulus is presented and how the response is recorded (Hazeltine, Ruthruff, and Remington, 2006). By paying attention to the modalities of the stimulus and response, the job relevance is preserved.

On the job, incumbents face stimuli in two key forms, audio – phone calls, and visual – emails. Their responses almost always involve a manual and visual task of entry into a computer system. Because this simulation uses these same modalities, the tool is highly relevant for most call center positions. The dual task costs observed in this setting are likely to be parallel to the dual task costs in the demanding multitasking environments on the call center floor.

Before the tool could be developed in its final form, several manipulations were checked and calibrated to ensure that the demands in the multitasking environment were realistic and yielded meaningful and practical variance. In a series of manipulation-check pilot studies with job incumbents, I tested different levels of many potential manipulations. These manipulations included the amount of time allotted to complete the task, the order of information presented in any given task, the length of the task itself, the frequency with which tasks were delivered, and the length of time allotted to open the next task. To determine the appropriate levels, I set up a series of trial tasks using a crude mock up of the simulation and several tape recorded stimulus examples. I observed performance on the different manipulated conditions and recorded performance with a stopwatch and accuracy checklist. Following the trial, I interviewed the pilot participants about the realism of the tasks and the demand of the environment. Through a process of trial and error, I was able to identify levels on each of the manipulated variables that pilot participants would typically endorse as realistic and appropriately demanding.

After the appropriate levels of these manipulations were identified, a request for proposals document was created and distributed to several software development firms. These firms had expertise only in computer programming; they were not testing vendors and had no meaningful expertise in the area of employment testing. The role of these vendors was simply to provide the computer programming expertise necessary to put the intellectual property of the host organization into a computer-delivered format for use with candidates.

While the multitasking simulation does measure performance in a highly relevant multitasking environment and the multitasking manipulations seem to drive performance levels, there are some weaknesses inherent in the tool that should be discussed. First, the tasks are much more complex than the tasks typically employed in laboratory multitasking research. The concern is that these complex tasks may be more "g" loaded than the simplistic tasks used in the lab setting. To combat this concern, the candidates are given detailed instructions, a learning session, and a practice session where the candidate gets to complete sample tasks without the key features of a multitasking environment present. The concern is also mitigated by the relevance of the nature of the simulation tasks with the nature of on-the-job tasks. That is, the task complexity is more demanding in this simulation than in most multitasking studies, but this complexity makes the simulation more generalizable to work related multitasking environments.

Second, task switching is often measured in terms of inspection time and reaction time. In this simulation, inspection time and reaction time cannot be separated. Rather the timing variables are likely a combination of both. Because the purpose of this study is not to understand how candidates process information in a multitasking situation, but rather to understand performance differences in a multitasking environment, distinguishing between these two types of task switching measures is not critical.

A third weakness of the multitasking simulation is that while every effort was made to keep the tool as job relevant as possible, the candidate's performance in the selection context is probably not the same as performance on the job. That is, in the selection setting candidates likely perform at their maximum level, which could be different than typical levels of performance across time on the job. Thus, the question of sustainability of performance in a multitasking environment over time cannot be answered completely with this 40-minute simulation tool.

One final weakness is that while every effort was made to eliminate confounding variables such as PC familiarity and typing skill. Some of these extraneous variables are still likely impacting performance on the simulation. To mitigate the concern about typing skill, I was careful about how typing tasks were evaluated in the simulation. There are very few typing requirements; only 26 of the 224 data points per candidate require typing into text fields. Further, when the candidate does have to type, s/he is scored based on whether or not s/he entered data into a text field, rather than the accuracy of this data entry.

PC familiarity can absolutely impact performance of the candidate on this simulation, but this was unavoidable. To mitigate this concern, every attempt possible was made to use only computer operations that are highly job related. The computer system used in the simulation is remarkably parallel to the systems actually used on the job, down to the smallest details. For example, I selected the color schemes to match as closely as possible the actual color schemes employed in the computer systems used on the job.

The data collected with this tool in an initial concurrent study, suggest that the outcome of the development process is a success. That is, the tool measures performance in terms of typical multitasking criteria in a job relevant and demanding multitasking environment. Data collected with this tool can be used to test several hypotheses to further both scientific and practical implications of applied multitasking research.

VITA

Theodore B. Kinney

EDUCATION

5/1998	B.A., Psychology The Pennsylvania State University University Park, Pennsylvania
8/2000	M.S., Industrial and Organizational Psychology Colorado State University Fort Collins, Colorado
8/2007	Ph. D., Industrial and Organizational Psychology The Pennsylvania State University University Park, Pennsylvania

PUBLICATIONS AND PRESENTATIONS

Kinney, T. B. (April, 2007). Individual Difference and Task Characteristic Predictors of Performance in a Multitasking Environment. Paper presented at the 22nd annual conference for the Society of Industrial and Organizational Psychology. New York, NY.

Murphy, K.R., Cleveland, J. N., Skattebo, A. L., & Kinney, T. B. (2004). Raters Who Pursue Different Goals Have Different Ratings. *Journal of Applied Psychology, 89(1),* 158 – 164.

Newman, D. A., Kinney, T. B., & Farr, J. L. (2004). Job Performance Ratings. In J. Thomas (Ed.). The Comprehensive Handbook of Psychological Assessment. New York, NY: Wiley and Sons.

Murphy, K. R., Cleveland, J. N., Kinney, T. B., Skattebo, A. L., Newman, D. A., & Sin, H. P. (2003). Unit climate, rater goals and performance ratings in an instructional setting. *Irish Journal of Management, 24*, 48-65.

Kinney, T. B. (2002). Contextual factors predict inter-rater agreement among Olympic judges. Paper presented at the 17th annual conference for the Society of Industrial and Organizational Psychology. Toronto, Canada.

Skattebo, A. L., Newman, D. A., Kinney, T. B., & Cleveland, J. N. (2002). Rater goals and unit climate affect rating behavior. Paper presented at the 17th annual conference for the Society of Industrial and Organizational Psychology. Toronto, Canada.