

REPORT DOCUMENTATION PAGEForm Approved
OMB No. 0704-0188

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing this collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number. PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.

1. REPORT DATE (DD-MM-YYYY) July 28, 2003		2. REPORT TYPE Final Report		3. DATES COVERED (From - To) Jan-July 2003	
4. TITLE AND SUBTITLE Measuring Multi-tasking Ability				5a. CONTRACT NUMBER N00014-03-M-0083	
				5b. GRANT NUMBER	
				5c. PROGRAM ELEMENT NUMBER	
6. AUTHOR(S) Susan C. Fischer, Kevin A. Morrin, and Susan Joslyn				5d. PROJECT NUMBER	
				5e. TASK NUMBER	
				5f. WORK UNIT NUMBER	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Anacapa Sciences, Inc. 301 E. Carrillo St. FL 2 Santa Barbara, CA 93101				8. PERFORMING ORGANIZATION REPORT NUMBER 1444-1	
9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES) Office of Naval Research Ballston Tower One 800 North Quincy Street Arlington, VA 22217-5660				10. SPONSOR/MONITOR'S ACRONYM(S)	
				11. SPONSOR/MONITOR'S REPORT NUMBER(S)	
12. DISTRIBUTION / AVAILABILITY STATEMENT Unclassified/Unlimited					
13. SUPPLEMENTARY NOTES					
14. ABSTRACT While multi-tasking (MT) may increase productivity, it also carries a tremendous downside including error, decreased morale, high training costs, high turnover rates, and attrition. This research shows that it is now possible to develop a test that would measure how individuals vary in their ability to concurrently perform multiple tasks under time limited conditions. The purposes of the present research were to (1) investigate complex real-world MT environments, (2) investigate existing measures of MT, and (3) begin development of a practical test of the ability. Current standards for educational and psychological tests served to guide the process of test development. Based on testing standards, a plan for development of an MT ability test was created. The initial phases of test development were also completed. The purpose, scope, and framework for the test are described in the report and the test specifications currently supported by empirical research are also given. This report also describes the additional research necessary for further development of a test of MT ability.					
15. SUBJECT TERMS Multi-tasking, ability, test, measures, LCAC, nursing, combat leader					
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT Unclass.	b. ABSTRACT Unclass.	c. THIS PAGE Unclass.	Unlimited	157 132	Susan C. Fischer
					19b. TELEPHONE NUMBER (include area code) 805-966-6157 Ext. 29

20030909 074

Measuring Multi-tasking Ability

Susan C. Fischer
Anacapa Sciences, Inc.

Kevin A. Morrin
Lock Haven University of Pennsylvania

Susan Joslyn
University of Washington

July, 2003

Prepared for:
Office of Naval Research
Dr. Susan Chipman

Ballston Tower One
800 North Quincy Street
Arlington, VA 22217-5660

ACKNOWLEDGEMENTS

This research has greatly benefited from the contributions of several individuals. We would first like to thank the project's program officer, Susan Chipman for her continuing pursuit of a robust test of multi-tasking ability. Science is only beginning to understand the ability to concurrently perform multiple tasks in complex real-world settings. A reliable and predictive measure of the individual differences that underlie performance in laboratory and real-world multi-tasking environments would greatly promote both scientific and applied interests. Susan Chipman's continued interest in multi-tasking has made this project a reality, and we are grateful to have had the opportunity to contribute to knowledge about this topic. We would also like to thank her for her insights about multi-tasking, directing us to important literature areas, and her able management of the project.

We would also like to thank the technical and support staff at Anacapa Sciences. We are grateful for the support we received from Amy Newsam, Barbara Gates, and Peggy Liborio Willis, who helped us assemble and format this report. We are also indebted to Bill Campsey and Jack Stuster, who were invaluable in helping us to arrange interviews with subject matter experts who work in multi-tasking jobs. Finally, we greatly appreciate the excellent review of this report provided by Doug Harris.

We have learned a great deal about multi-tasking in this research. Most of what we have learned has been taught to us by the individuals who perform jobs that require this ability. Each of the jobs that we investigated demand an almost unbelievable degree of multi-tasking skill. We were truly amazed at what these individuals accomplish in their daily work. We would like to thank the ICU and floor nurses, chefs, Army combat leaders, LCAC Craftmasters and LCAC Navigators who educated us about multi-tasking in the working world.

EXECUTIVE SUMMARY

INTRODUCTION

We live and work in a world that frequently requires the performance of multiple tasks within a limited time period, requiring a capability that has become known as multi-tasking (MT). While MT may not be present in everything that we do, it is getting more difficult to find work environments in which MT is not at least part of the job. Both military and civilian work environments require MT. For example, the crew aboard the Navy's Landing Craft Air Cushion (LCAC) is tremendously busy performing multiple tasks within a short period of time. Nurses, air traffic controllers, and chefs are examples of civilian positions that place heavy demands on MT ability.

While multi-tasking may increase productivity and reduce overall costs, it also carries a tremendous downside. The negative consequences of MT come in several forms, one of which is increased probability of error. When the human information processing system is used to capacity, as is often the case when multi-tasking, a likely outcome will be error. Unfortunately, human error in decision-making under time-limited situations has been the cause of several disasters in each of these types of jobs. The air collision in German airspace in 2002 that was the result of air traffic control (ATC) error is only one example.

Another negative consequence of MT in the workplace is decreased morale, which nearly always leads to high levels of burnout, turnover rates, and attrition. MT is, by its very nature, stressful. Hence, many jobs that require MT also have high turnover rates and attrition. These jobs often require extensive training, and organizations invest a great deal of money to train selected applicants only to lose them later because of the stressful nature of the work.

MT not only increases the probability of error, burnout, stress, attrition, and training costs. Every time an individual switches to another task, it takes a small amount of time to reorient to the new task. While it may seem like productivity is increased by reducing staff and increasing task load, overall performance may actually be slowed by MT.

Despite the problems associated with MT, not every air traffic controller or nurse experiences stress, burnout, or makes a large number of errors. Some individuals seem resistant to the negative effects of MT, and even seem to thrive on the challenge. Some individuals are much more able to perform well in multi-tasking environments than others. In psychological terms, there may be a general ability to concurrently organize and perform more than one task, which allows some people to perform well in MT environments.

Recent research supports this hypothesis, showing that normal adults vary in how well they perform laboratory tasks requiring the simultaneous performance of multiple tasks under time-limited conditions (Joslyn & Hunt, 1998). What is even more impressive is that an abstract laboratory task used in this research predicts ultimate performance on laboratory simulations of jobs that require multi-tasking (emergency dispatching, emergency call answers, and air traffic control (ATC)) (Joslyn & Hunt, 1998).

If individuals truly vary in their ability to multi-task, it should be possible to measure that ability and use the assessment to predict future performance in MT environments. In other words, it should be possible to develop a measurement instrument (a test) that could be used to screen individuals for positions that demand high levels of MT ability. Joslyn's and Hunt's work strongly suggests that development of such a test is possible. Indeed, their laboratory task, the Abstract Decision-Making (ADM) task may be a direct measure of MT ability.

A test that could reliably measure MT ability and could predict job performance in a variety of MT environments would be highly useful. Training costs for many MT jobs could be reduced by using the test to select those individuals who would perform well on the job. However, this report will also show that, while previous research has produced a great deal of knowledge about MT in relatively simple, controlled, laboratory settings, little is known about MT in complex real-world environments. To create a reliable and valid predictor of MT ability in real-world settings, a better understanding of complex environments is needed. The similarities and differences among MT environments have not been studied. As a result, we do not yet understand the kind of real-world performance a test of MT ability should predict. Moreover, there are no existing tests of MT ability for normal populations. The literature does include various laboratory tasks and paradigms that might form the basis of a future test of MT. However, usable tests have not been developed.

PURPOSE OF THIS RESEARCH

The purpose of the present research was twofold. The first purpose was to begin to close the gap in our knowledge of real-world MT. Second, this research also began the process of developing a usable and practical test of MT ability. A two-pronged approach was taken to better understand (1) complex MT environments and (2) existing measures of MT. Four MT environments were studied to begin to understand the cognitive operations they demand. A preliminary ontology of cognitive operations required by MT was developed and used to analyze the environments. The results of the analysis of MT environments were used to establish preliminary requirements for a predictive test of performance in those settings.

A review of the literature was also conducted to (1) identify current measures that could potentially be used to predict MT performance in real world settings, and (2) analyze those measures to determine the kinds of cognitive operations they measure. To begin the process of developing a usable and practical test of MT ability, current standards for educational and psychological tests were studied. Based on four phases of test development prescribed by the standards, a plan for development of an MT ability test was created. Following the plan, the initial phases of test development were completed. This report also describes the additional research necessary for further development of a test of MT ability. A set of studies was designed to lay the requisite empirical groundwork for test development and to examine the construct and predictive validity of the resulting test. These studies are fully described in Chapter Six of this report. In this executive summary, we provide an overview of the findings of this study.

INVESTIGATION OF MULTI-TASKING ENVIRONMENTS

Generally stated, the purpose of investigating MT environments was to gain knowledge about the kind of performance a test of MT ability should predict. This research constitutes an initial examination of the criterion performance the proposed test seeks to predict. A better understanding of the similarities and differences among MT environments is imperative to development of a test that can predict performance in a wide variety of MT environments.

Several issues were important to this study. First, how similar and how variable are MT environments in terms of the kinds of cognitive requirements they place on individuals who work in them? Do they all require the capacity to remember lots of information, for example? Do they all require the interleaving of tasks, and hence the ability to use prospective memory? Is the ability to prioritize important to all MT environments? Which cognitive capabilities make someone good at MT jobs?

METHOD

Participants. Based on several criteria, two military MT environments were selected for study: operation of the Navy's Landing Craft Air Cushion (LCAC) and Army combat unit command. Both the Craftmaster and the Navigator positions aboard the LCAC were investigated. Two civilian MT environments were also selected for study: restaurant food preparation/chef and nursing. Both of these civilian environments experience high turnover rates and financial losses in training costs due to burnout. Nine professionals who worked in four different MT environments participated in the interviews. Each of the participants was highly experienced and qualified in their own field.

Materials. A standard set of questions was designed to probe the cognitive requirements of work environments, regardless of the particular field of work or job content. The questions were designed for use in the context of a critical incident of MT that the participant had experienced as part of his or her work.

Procedure. Interviews were conducted with each participant. The interviewer first described the purpose of the study and asked questions about the participant's qualifications and background. Then participants were asked to describe a critical incident that they had experienced in their work. They were asked to think of an incident that heavily demanded MT performance. After describing the incident, the interviewer asked a series of questions pertaining to six different topics related to the cognitive requirements of the job including issues of memory, task prioritization, decision-making, knowledge and experience, the work environment, and relationships among the components tasks.

RESULTS

Each of the four MT environments are described in detail in this report. The results indicated the each environment possesses eleven characteristics of MT settings originally specified by Burgess (2000) and further elaborated in this report.

The results also indicated that the jobs varied somewhat in the kinds of cognitive operations they required. The memory requirements they place on workers were very

similar. All of the jobs require required STM storage of information (e.g., headings for LCAC navigators and operators, vital signs for nurses). LTM retrieval of domain-specific knowledge learned in training or on-the-job experience was also necessary in each of the jobs we studied. Most, but not all, jobs required prospective memory. Updating of working memory was extremely important to all of the jobs. The need to maintain situation awareness, whether one is a combat leader, nurse, chef, or LCAC crewmember, is critical in these dynamic environments.

The control of attention was also critical to performance. In each environment, multiple sources of information were available and were often presented simultaneously. For this reason, workers must decide whether to selectively focus on one piece of information, or divide their attention among several. The relative importance of information seems to be the key determinant whether one takes the strategy of dividing or focusing attention. If the consequences of missing information are severe, one must use a divided attention strategy. All jobs required both selective and divided attention.

The fact that multiple, very different tasks are required by these environments means that (1) workers must switch mental sets when going between tasks and (2) that prioritizing is key to good performance. Indeed, each of our respondents, with the exception of the LCAC operator, reported that prioritization was key. They also reported that it was the hardest element of the job, and took them the longest to learn. Based on their responses, if there is one factor that determines whether one does well in these jobs or not, it is the ability to prioritize effectively.

CONCLUSIONS

If we were to design a test of MT ability that would incorporate the cognitive operations most real-world MT environments require, what would it include? Based on the results of our analysis, we propose a test should require that test takers engage in the following cognitive operations.

- STM memory storage
- LTM retrieval
- Prospective memory
- WM updating and monitoring
- Mental Set Switching
- Classification
- STM rehearsal
- Control of attention required by simultaneous presentation of stimuli
- Prioritization

ANALYSIS OF EXISTING MEASURES OF MULTI-TASKING

To better understand the cognitive processes and operations that current measures of MT assess, we first conducted a thorough review of the literature to identify measures that other researchers have used.

METHOD

Relevant literatures residing on a variety of databases were searched. The resulting hits were examined for relevance and high payoff sources were obtained. Selected sources were reviewed and pertinent information was extracted about measures of MT. A systematic search of the most recent (within the past 5 years) relevant literature was conducted in which a variety of academic and government databases was queried.

RESULTS

Researchers have studied MT using various types of measures. One type has been employed to assess neuropsychological disorders; measures involved the application of strategy, planning, and executive control of working memory. These measures include the Multiple Errands Test (MET), the Six Element Test (SET) and the Greenwich test. A second type has been employed in the simulation of work environments. These measures include SYNWORK, the Multiple Attributes Test Battery (MATB) and the Abstract Decision Making task (ADM). A third type, stemming from basic research efforts, has addressed the limitations of human performance. Here, the dual- or tri- task paradigm has been used to assess how individuals distribute cognitive, perceptual, and motor resources in laboratory situations that contain multiple simultaneous demands. Information coordination tasks and the psychological refractory period (PRP) paradigms have also been used. The cognitive operations that the measures incorporate are discussed in detail in this report. The conclusions we draw from these results are given in the next section of this executive summary.

GAPS IN THE MEASUREMENT OF MT

Laboratory tasks (including the dual task paradigm, information coordination tasks, and the psychological refractory period procedure), which have been extensively and successfully used to examine the fundamental limits of cognition, do not adequately represent the complexity of real-world MT environments in terms of the cognitive operations they demand. First, they typically do not require prospective memory, which is critical to successful performance in the real-world MT jobs we analyzed. Second, while many of the jobs we analyzed required the continuous storage of information in STM, STM rehearsal, and LTM retrieval, these elemental tasks place little demand on these forms of memory and instead rely on iconic or auditory storage. Third, they do not assess more important complex and demanding cognitive processes used in real-world MT environments such as planning and deductive logic. Finally, while these MT measures do require the participant to prioritize among tasks, we believe that they demand only the simplest kind of prioritization, which does not adequately represent the complexity of real-world MT environments.

The measures that have been developed to assess neurological problems, such as dysexecutive disorder, also fail to adequately represent the cognitive components of MT jobs. While the MET, SET, and Greenwich tests do assess cognitive operations such as setting and following a plan, retrieving information from LTM, storing and using information in STM, remembering future tasks (prospective), and switching among different tasks, they do not present a situation in which a person must divide attention among *simultaneously* presented multiple sources of information nor do they require

selective attention. In each of the jobs we analyzed, the need to divide or select attention was a salient and critical component of the environment. Indeed, it is part of what creates an MT environment as the worker cannot control when he or she will receive information. As is true of basic laboratory tasks, these neuropsychological measures also do not represent the complexity of prioritization and deductive logic found in real-world MT jobs. Moreover, it is highly likely that ceiling effects, or at least range restrictions, would be found in normal populations who take the MET, SET, and Greenwich tests.

Perhaps it is not surprising that the tests that have been purposely designed to simulate or predict performance in real-world jobs appear to best represent the cognitive operations we believe those jobs demand. The MATB could not be used as the basis of a general test of MT ability because its content is taken from aviation. The other two measures in this category (SYNWORK and ADM), however, are good candidates on which to base an MT ability test. If choosing between SYNWORK and ADM the immediate obvious choice would be ADM if for no other reason than it has already been demonstrated to predict simulated and actual job performance at a surprisingly high level of accuracy. This empirical reality is no small consideration as it is highly unusual to obtain the level of predictive power that has been demonstrated with ADM. There is no real need to consider the capabilities of SYNWORK given this advantage of ADM. However, there are other compelling reasons to base a test of MT ability on ADM, which are thoroughly discussed in this report.

In conclusion, current knowledge of MT and its measurement strongly suggest that the best candidate for predicting MT ability is ADM. The goal to develop an assessment test of MT ability would be best reached by basing the test on ADM. However, it should also be recognized that it is premature to conclude that ADM will predict performance in all, or even most, MT environments. ADM has successfully predicted reliable performance measures of dispatching and ATC. But it may be that these particular jobs shares specific characteristics not found in other jobs.

There are additional issues surrounding the use of ADM as a predictor that also must be addressed. Perhaps the most important cognitive skill that is not adequately assessed by ADM is the ability to prioritize. This issue and others concerning ADM as a test of MT ability are thoroughly discussed in this report.

DEVELOPMENT OF AN MT ABILITY TEST

Significant progress was made in the present research toward the development of a test of MT ability. Although full development of the test is beyond the scope of the current project, the initial phases of design have been completed. To ensure that the proposed test of MT ability meets criteria recognized by the scientific, educational, and testing communities, design was guided by current testing standards published jointly by the American Educational Research Association (AERA), American Psychological Association (APA), and the National Council on Measurement in Education (NCME) (1999). Using the standards to guide the process of test development and evaluation also ensures the MT test (1) will be of the highest quality, (2) can be safely used by government agencies and private industries, and (3) can be commercialized. Finally, the

standards provide a framework on which to organize and evaluate the development process.

Development of an MT ability test was approached with careful consideration of current standards (AERA, APA, NCME, 1999). The AERA et al. (1999) document prescribes and describes a four-phase approach to test development and provides enumerated criteria that all educational and psychological tests must meet.

The MT test will be based on Joslyn's and Hunt's (1998) ADM. Full development of a test that would meet current standards (AERA, APA, NCME, 1999), however, requires additional research. Previous research and the present study provide a sufficient understanding of MT, as an ability and psychological construct, to specify the purpose and scope of the test. A framework for the test can be developed at this point, which should describe the extent of the domain to be assessed and the scope of the construct (AERA, APA, NCME, 1999).

Additional research, however, is necessary to complete the second phase of test development, which requires test design to be taken to a higher level of specification. As previously discussed, the first phase of test development focuses on establishing clear definitions of the proposed test's purpose and scope. A framework for the test is developed that extends the purpose of the test to describe the construct to be measured. The framework delineates aspects of the construct that are targeted by the test. What follows documents the intended purpose, scope, and framework for a test of MT ability.

PURPOSE

The MT test will serve a scientific measurement purpose that can be practically used to address applied needs in MT environments. Broadly stated, the purpose of the test will be to measure individual differences, within normal populations, in multi-tasking ability. In so doing, the test can be used to identify those individuals who are likely to perform well in environments or jobs that require high levels of MT ability. The test will incorporate a scoring system that predicts measures of asymptotic performance in real-world MT environments, as well as measures of time required to reach asymptotic levels. Hence, it will be both a test of ultimate performance and a test of skill acquisition.

MT ability is a psychological construct that has received increasing attention in the basic and applied literature. Simple stated, the MT construct is the ability to concurrently perform or interleave multiple tasks. MT ability is thought to place heavy demands on several executive control functions, which many theoretical accounts include as part of working memory. Despite its probable overlap with the working memory construct, current findings indicate that MT ability is a distinct individual difference variable. Current findings also indicate that it has little to no relationship to other constructs such as processing speed and fluid intelligence. These conclusions, however, warrant further investigation. MT ability also incorporates the ability to prioritize the many tasks that must be performed. A body of research exists that supports the existence of individual differences in the ability to concurrently perform or interleave multiple tasks. Recent research has succeeded in measuring such differences and predicting performance in real-world environments and jobs that require individu-

als to use the ability. The test will be based on a recently developed laboratory task of time-pressured decision-making that has been shown to be highly predictive of simulated emergency dispatching and ATC job performance.

SCOPE

The test is intended to discriminate differences in MT ability among normal populations of adults. Although a body of research has associated MT ability with dysexecutive syndrome and a variety of other neuropsychological disorders that involve impairment of executive control functions, the test is not intended as an instrument to diagnose or otherwise measure such disabilities. The test is intended for adult populations who work in real-world MT environments, and should not be used to discriminate differences among children or aged populations. The test is also intended to have limited criterion validity with respect to work environments. It is intended to predict relevant measures of performance in MT environments, but not in stressful, fast paced, nor time-limited environments; however similar these environments may be to MT jobs.

FRAMEWORK

The present research provides a logical framework for understanding MT ability and the proposed MT ability test. Standards recognize that this framework may change as test development proceeds through the interplay between construct development and test development. However, current analysis supports basing the MT ability test on the cognitive requirements commonly found in real-world MT jobs. Hence, the MT ability test will incorporate cognitive operations that current analysis shows are critical to successful MT performance. The cognitive operations that appear to be critical are STM rehearsal and storage, WM updating, prospective memory, divided attention, selective attention, mental set switching, LTM retrieval, and prioritization.

Analysis of the ADM task reveals that its current version incorporates and requires participants to employ a set of cognitive operations that are a good match to the operations required by MT environments. Short-term, prospective and working memory operations are integral to both ADM. Executive control functions such as mental set switching, selective attention, divided attention, and rehearsal for STM are also required by ADM.

The ability to effectively prioritize multiple tasks appears to be a critical function that workers must perform in MT environments. While the ability to effectively prioritize multiple tasks in the real world is what makes or breaks a worker, however, we currently do not know if ADM can be performed relatively successfully without this skill. However, it may be possible to increase the degree to which ADM measures the ability to prioritize tasks by modifying ADM's structure, scoring system, or rules. The importance of prioritization to real-world performance in MT jobs warrants investigation of modifications to ADM to better represent the ability to effectively perform this operation.

ADM also fails to incorporate a LTM retrieval component in the sense that domain-specific declarative or procedural knowledge that is typically learned through extensive on-the-job experience is not utilized in ADM. However, any abstract test that would be

applicable to many job domains would necessarily not include LTM retrieval in the way it is used in real-world environments. The requirement that the test be applicable to a wide variety of jobs appears to preclude any meaningful LTM retrieval component. Hence, current and modified versions of the ADM task will be designed to measure eight critical cognitive components required by MT environments, which include STM rehearsal and storage, WM updating, prospective memory, divided attention, selective attention, mental set switching, and prioritization.

DEVELOPMENT AND VALIDATION OF MT TEST

Research questions that are particularly important to development of a test of MT ability are identified in this report. Issues that must be resolved in developing ADM as a test include test length, response format, test difficulty, feedback, instructions and test administration, the role of prioritization, and the role of deductive logic. To address these issues, seven studies were designed. Several of the studies serve the purpose of resolving issues pertaining to test design and assembly. The last two studies address issues regarding MT as a construct and validation of the MT ability test. Each of the issues refers to requirements that are incorporated in the third and fourth phases of test development as prescribed by the AERA et al. (1999) standards.

A hierarchical relationship is evident among the issues. Questions most pertinent to test development (test length, response format, test difficulty, instructions, administration, cognitive components, feedback) must be addressed before psychometric properties (reliability, construct validity, and predictive validity) may be estimated. The purpose of each of the seven studies is given below.

STUDY #1: TEST ADMINISTRATION PROCEDURES AND INSTRUCTION

PURPOSE. The primary purpose of the first study will be to assess the effects of changes to ADM in terms of test administration procedures and instructions.

STUDY #2: RESPONSE FORMAT

PURPOSE. The primary purpose of the second study will be to examine issues of response format.

STUDY #3: FEEDBACK STUDY

PURPOSE. The primary purpose of the of this study will be to examine how changes in the kind and amount of feedback provided in ADM affect its ability to predict performance in simulated or real MT environments.

STUDY #4: PRIORITIZATION

PURPOSE. The primary purpose of the fourth study will be to examine how changes to the structure of ADM to include a greater emphasis on prioritization will affect its ability to predict performance in simulated or real MT environments. Having established the basic features of the MT ability test, in terms of response format and feedback, we will begin to examine issues concerning the cognitive operations the test requires.

STUDY #5: DEDUCTIVE LOGIC DEMAND (BIN OVERLAP)

PURPOSE. The primary purpose of the fifth study will be to examine how changes to ADM's requirement for deductive logic affects its ability to predict performance in simulated or real MT environments.

STUDY #6: CONSTRUCT VALIDITY

PURPOSE. The first five studies have been designed to ferret out issues concerned in test development. Study #6 is the first study to be conducted on a completely designed test. The primary purpose of the sixth study will be to examine the test's construct validity. This study will attempt to resolve questions concerning the relationship of MT ability to other constructs. Is MT a separable construct? Alternatively, is MT ability a component of WM, processing speed, or fluid intelligence. Several models will be developed and evaluated using latent variable analysis.

STUDY #7: PSYCHOMETRIC PROPERTIES

PURPOSE. The primary purpose of the final study will be to examine the psychometric properties of the final version of the test. At this point in time, the research will have produced a completed test. The relationships between the individual differences measured by the new MT ability test and those of other constructs will have been examined in Study #6. It will now be important to establish the degree to which the new version can predict performance in other MT environments. It is important to note that the test development process has, in fact, ensured that the MT test has predictive capability. At every step along the way, the criterion for decisions about test development were based on which version predicted a simulation of 911 dispatching. The consistent use of emergency dispatching simulation provides a necessary stable base of comparison. Attempts to use other measures of performance in other MT environments would only confuse the test development process. However, consistent use of the emergency dispatching simulation also limits the criterion validity of the test. In study #7, this limitation will be evaluated.

CONCLUSIONS

The research described in this report broadens and deepens current knowledge of real-world MT. It makes significant contributions to the study of MT. The research provides a way to define MT environments that was previously unavailable to researchers. Comparison of four MT settings and 8 different jobs in those settings showed that although MT environments appear to differ greatly, they share a number of characteristics. The definition of MT environments has also afforded a path by which cognitive operations that might be demanded by these environments can be specified. The cognitive operations have been used in this research to illuminate important aspects of MT. For example, some appear to be more important to MT environments than others. Several appear to characterize complex MT environments from simple ones.

This research has also provided a way to identify requirements for a test of MT. A test of MT ability is not yet available to researchers. Because measurement forms the basis of all research, development of a test would greatly advance researchers ability to

study MT. The present research lays the groundwork for measurement of MT to begin. Initial test design has been completed according to standards and a series of studies necessary to further test development and evaluation have been designed. Future research that addresses the research issues discussed in this report will produce a greater understanding of what is now a very common activity in our world.

CONTENTS

	Page
Chapter One: Introduction	1
The Downside to MT	2
MT as a Measurable Ability	3
Purpose of this Research	5
Organization of Report.....	5
Overview of Literature Related to Multi-Tasking	6
Working Memory Capacity	7
Information Coordination.....	12
Personality Factors	14
Chapter Two: Technical Research Objectives.....	16
Chapter Three: Investigation of Multi-Tasking Environments	17
Selection of MT Environments	19
Defining MT Environments	19
Selected MT Environments	22
Interviews	24
Participants.....	24
Interview Questions and Techniques.....	25
Interview Procedures.....	25
An Ontology of Cognitive Operations Used in MT	26
Interview Results.....	28
Description of LCAC Environment	28
LCAC Environmental Characteristics	32
LCAC Cognitive Operations	32
Description of Army Combat Command Environment.....	35
Army Combat Environmental Characteristics.....	37
Army Combat Command Cognitive Operations	38
Nursing Environment.....	42
Nursing Environmental Characteristics	44
Nursing Cognitive Operations	44
Restaurant Kitchen Environment.....	46
Restaurant Kitchen Environmental Characteristics	47
Chef Cognitive Operations	48
MT Environment Analysis Conclusions	49
Environmental Characteristics	49
Cognitive Operations.....	51
Conclusions	54

Chapter Four: Analysis of Existing Measures of Multi-Tasking	54
Literature Review Methods	54
Results	55
Measures Designed to Assess Neuropsychological Disorders.....	55
Measures Designed as Simulations of Real-World Work Environments	59

CONTENTS (CONTINUED)

	Page
Measures Designed to Investigate the Limits of Human Information Processing.....	65
Chapter Five: Gaps in the Measurement of MT	71
Comparison of MT Measures	71
Additional Issues Surrounding ADM as an MT Ability Test	74
Planning.....	74
Prioritizing.....	75
Prioritization and the Current Version of ADM.....	78
The Issue of Feedback and Prioritization in ADM	79
The Issue of Structural Relationships Among Tasks in ADM	80
Summary	
Chapter Six: Development of an MT Ability Test.....	82
Test Standards.....	82
Overview of MT Test Development	84
Test Purpose, Scope, and Framework: Phase I	85
Purpose	85
Scope	86
Framework	86
Test Specifications: Phase II (Standard 3.3)	87
Test Taker Populations	88
Content and Difficulty of Test: Discrimination (Standard 1.6).....	88
Item Formats	89
Psychometric Properties of Test and Items	89
Item Arrangement.....	90
Number of Items.....	90
Time for Testing.....	90
Directions to Test Takers	91
Procedures for Test Administration (Standard 3.6)	91
Procedures for Scoring (Standard 3.6).....	91
Response Format (Standards 1.7, 3.6).....	91
Norm Referenced or Criterion Referenced.....	92
Assembly, Field Testing, and Evaluations: Phases III and IV	92

Chapter Seven: Development and Validation of MT Test.....	92
Issues of Test Design and Assembly	92
Test Length.....	92
Response Format	93
Test Difficulty.....	93
Feedback	93
Instructions and Test Administration (Standard 2.8)	94
Emphasis on Prioritization	94
Emphasis on Deductive Logic.....	94

CONTENTS (CONTINUED)

	Page
Issues Concerning MT as a Psychological Construct (Standard 1.8).....	95
Issues Concerning Psychometric Attributes of Test.....	95
Reliability.....	96
Criterion Validity	96
Norms.....	96
Seven Studies to Develop and Validate Test.....	96
Study #1: Test Administration Procedures and Instruction.....	96
Study #2: Response Format.....	98
Study #3: Feedback Study	99
Study #4: Prioritization.....	100
Study #5: Deductive Logic Demand (Bin Overlap).....	101
Study #6: Construct Validity.....	102
Study #7: Psychometric Properties	104
Chapter Eight: Conclusions.....	106
References	107
Appendix A: Sources Reviewed	115
Appendix B: Summary of Standards	121

LIST OF FIGURES

Figure	Page
1 Research strategy to design test	17
2 Hierarchical model where MT is a component of gF.....	104
3 Model where MT is entirely separable from other constructs.....	104

LIST OF TABLES

Table	Page
1 Defining Characteristics of MT Environment	21
2 Candidate MT Jobs.....	23
3 Summary of Environmental Characteristics of MT Environments	50
4 Summary of Cognitive Operations Required by MT Environments.....	52
5 Summary of Cognitive Operations of Existing Measures of MT	75

MEASURING MULTI-TASKING ABILITY

CHAPTER ONE: INTRODUCTION

We live and work in a world that frequently requires the performance of multiple tasks within a limited time period, requiring a capability that has become known as multi-tasking (MT).¹ While MT may not be present in everything that we do, it is getting more difficult to find work environments in which MT is not at least part of the job. Nursing, nuclear power control room operation, emergency medicine, emergency dispatching, air traffic control, and mid-level management are just a few modern civilian jobs that demand MT. Within each service of the military, it is even easier to find examples. Operators of the Command Information Center (CIC) aboard Navy ships, crewmembers of the Landing Craft Air Cushion (LCAC), aircraft pilots, and the bridge officer aboard an aircraft carrier are just a few Navy jobs that require MT. The Air Force and Army also demand MT of their personnel as pilots, leaders of combat units, and decision-makers responsible for distributing resources (e.g., artillery) on the battlefield. Observation of an Army Tactical Operations Center (TOC) would provide a quintessential picture of MT activity.

MT is not limited to work environments as it is increasingly found in mundane settings as well. Cell phones now interrupt the flow of activity in ordinary places such as grocery stores, vehicles, restaurants, and movie theatres. The interruptions they create encourage the interleaving, or simultaneous execution, of tasks. Internet, television, and radio communications make available vast amounts of information that can be processed while performing other tasks like paying bills, riding a stationary bicycle, or cooking dinner. Automated appliances provide the ability to wash clothes, clean dishes, answer the phone, and cook dinner all at the same time.

Perhaps technological, economic, and sociological forces have combined to effect an increase in MT activity. For example, technologies now give military command centers complete information about placement of friendly forces, which has created the additional task of forming a coherent understanding of the situation from an overload of stimulus material. Communication technologies such as faxes, mobile phones, and pagers are widely available now, which affords immediate communication as well as unpredictable interruption of other tasks. Economic forces are perhaps the greater source of increase in MT activity. Higher demands on productivity in the workplace mean that more tasks must be completed in a given period of time. With productivity, cost cutting, and efficiency as their objectives, private and government organizations have reduced staffing without a corresponding reduction in workload. For example, each of the services has experienced downsizing in the past 10 years without a concomitant reduction in requirements. The Navy has reduced staffing aboard some ships from 395 sailors to 95. As a result, the work that used to be performed by several must

¹ The technical definition of MT and the characteristics of MT environments are addressed in Chapter 3 of this report.

now be performed by one. The same trends have been observed in civilian industries. For example, the high costs of medical care have motivated hospitals to cut costs by reducing nursing staffs. Hence, nurses are typically required to carry a higher patient load today than they were in previous years.

Perhaps we have also moved toward increasing MT activity in the workplace and at home because of sociological factors pertaining to social structures and values. For example, telecommuting, job-sharing, and families' attempts to decrease the amount of time spent at work are all sociological factors that have increased how much MT is carried out at home and at work. In short, we are fitting in more tasks in any given unit of time in all aspects of life.

THE DOWNSIDE TO MT

While multi-tasking may increase productivity and reduce overall costs, it also carries a tremendous downside. The negative consequences of MT come in several forms, one of which is increased probability of error. When the human information processing system is used to capacity, as is often the case when multi-tasking, a likely outcome will be error. Indeed, this is exactly what has been observed in medicine, power plant operation, piloting, and air traffic control. In many cases, errors go unnoticed, are corrected before consequence, or do not produce severe consequences. For example, research has shown that nurses make errors in fluid medication delivery on a weekly basis (Fischer & Harp, 1999). Usually, a patient receives a little more or a little less medication or their treatment is delayed. However, the potential for a serious error is always present in nursing, and is realized at frequencies that health organizations are only beginning to monitor. The reality is that many of the jobs that place heavy MT demands on personnel have the potential to result in disaster. For example, consider the potentially disastrous consequences of error for CIC operators, platoon leaders in a fire-fight, nurses, nuclear control room operators, pilots, air traffic controllers, and emergency medical technicians. Unfortunately, human error in decision-making under time-limited situations has been the cause of several disasters in each of these types of jobs. The air collision in German airspace in 2002 that was the result of air traffic control error is only one example.

Another negative consequence of MT in the workplace is decreased morale, which nearly always leads to high levels of burnout, turnover rates, and attrition. MT is, by its very nature, stressful. When time is limited, tasks are many, and the consequences are high, stress and burnout are extremely likely. Hence, many jobs that require MT also have high turnover rates and attrition. Three of the four MT environments studied in the present research (nursing, LCAC Navigation and Operation, and restaurant food preparation) are burdened by high attrition rates. The present research revealed an attrition rate of 70% for LCAC Navigators! High turnover is extremely costly. These jobs often require extensive training, and organizations invest a great deal of money to train selected applicants only to lose them later because of the stressful nature of the work.

MT not only increases the probability of error, burnout, stress, attrition, and training costs. Research suggests that the ostensible benefits of MT may be illusory. Several

researchers (Pashler, Johnston, and Ruthruff, 2001; Rubinstein, Meyer & Evans, 2001, among others) have argued that switching among multiple tasks produces performance deficits compared to single task conditions or blocked trials. Every time an individual switches to another task, it takes a small amount of time to reorient to the new task. Hence, task switching is associated with slower performance times. While it may seem like productivity is increased by reducing staff and increasing task load, overall performance may actually be slowed by MT.

MT AS A MEASURABLE ABILITY

Despite the problems associated with MT, not every air traffic controller or nurse experiences stress, burnout, or makes a large number of errors. Some individuals seem resistant to the negative effects of MT, and even seem to thrive on the challenge. Although the requirement to switch among tasks may slow performance for everyone (Pashler, Johnston, and Ruthruff, 2001, Rubinstein, Meyer & Evans, 2001), the degree of that decrement may vary among individuals. Some individuals may use strategies that afford efficient prioritization of tasks resulting in superior performance in MT environments. Or, individual differences in personality variables, such as risk taking and confidence, may positively influence willingness to engage in MT and thereby increase performance. In short, some individuals are much more able to perform well in multi-tasking environments than others. In psychological terms, there may be a general ability to concurrently organize and perform more than one task, which allows some people to perform well in MT environments.

Recent research supports this hypothesis, showing that normal adults vary in how well they perform laboratory tasks requiring the simultaneous performance of multiple tasks under time-limited conditions (Joslyn & Hunt, 1998). What is even more impressive is that an abstract laboratory task used in this research predicts ultimate performance on laboratory simulations of jobs that require multi-tasking (emergency dispatching, emergency call answers, and air traffic control (ATC)) (Joslyn & Hunt, 1998).

If individuals truly vary in their ability to multi-task, it should be possible to measure that ability and use the assessment to predict future performance in MT environments. In other words, it should be possible to develop a measurement instrument (a test) that could be used to screen individuals for positions that demand high levels of MT ability. Joslyn's and Hunt's work strongly suggests that development of such a test is possible. Indeed, their laboratory task, the Abstract Decision-Making (ADM) task may be a direct measure of MT ability.

A test that could reliably measure MT ability and could predict job performance in a variety of MT environments would be highly useful. It could be used by many different civilian and military organizations to discriminate individuals who are likely to perform well in MT jobs from those that will probably perform poorly. Nursing schools could use such a test as a counseling tool to guide their students to work environments appropriate to their abilities. Law enforcement agencies could use it to identify applicants who would likely do well in emergency dispatching jobs. The Navy could use it to screen the very large and heterogeneous pool of LCAC Navigator applicants, as well as applicants for many other Navy jobs that demand high levels of MT. The Army could

use it to assist Army officers in identifying their strengths and weaknesses so as to guide their personal leadership development programs. Training costs for many MT jobs could be reduced by using the test to select those individuals who would perform well on the job. There is certainly a need for a test that could reliably measure MT ability and predict job performance in MT environments. There is also considerable demand for such a test. Since the present research began, the authors of this report have received numerous messages, letters, and phone calls from a variety of sources requesting just such a test.

However, the utility of an MT ability test would depend on many factors. First, it would be necessary to firmly ground all aspects of the test in documented empirical findings. Its psychometric properties would have to be superior and demonstrated to the scientific community. Research would have to show that the test (1) was a stable measure, (2) measured MT ability and not other constructs for which tests are available, (3) predicted performance in several MT jobs, and (4) met current standards for psychological tests.

The scientific community has attained a level of knowledge about MT where it may now be possible to reliably measure MT ability. Hence, there is substantial promise that an acceptable test could be developed. This report will show that several laboratory tasks have been developed that might be used as the basis for such a test. There is particular promise with Joslyn's and Hunt's (1998) ADM task as it has already been shown to predict several measures of performance for jobs that demand MT.

However, this report will also show that, while previous research has produced a great deal of knowledge about MT in relatively simple, controlled, laboratory settings, little is known about MT in complex real-world environments. Numerous laboratory tasks have been used to investigate the limits of performance under multiple task conditions. However, very few investigations have studied MT as it occurs in real-world settings. Meyer and Kieras (1997) provide an excellent review of the substantial knowledge garnered over the past 40 + years about MT in simple controlled environments. The present research will show, however, that real-world MT environments are far more complex and most laboratory task paradigms are far too simple to use as predictors of MT in complex environments.

To create a reliable and valid predictor of MT ability in real-world settings, a better understanding of complex environments is needed. The similarities and differences among MT environments have not been studied. As a result, we do not yet understand the kind of real-world performance a test of MT ability should predict. We know little about how the ability develops, or does not develop, with experience on the job. Applied research on MT has been largely limited to populations with neuropsychological disorders (e.g. Burgess, Veitch, de Lacy Costello & Shallice, 2000; Shallice & Burgess, 1991). Although several tests have been developed to diagnose patients with neuropsychological disorders related to MT, no tests of MT ability have been developed for normal populations. Hence, we know little about the cognitive processes that existing measures of MT presumably tap. In fact, Joslyn and Hunt (1998) conducted the only set of studies that have attempted to determine if MT is a separable ability from other potentially related psychological constructs such as working memory (WM), short term

memory (STM), fluid intelligence, and processing speed. While research supports the hope that a reliable and predictive test of MT could now be developed, additional research is needed to better understand MT ability as a construct, MT real-world environments, and existing measures of MT.

PURPOSE OF THIS RESEARCH

The purpose of the present research was to investigate some of the issues noted above and to begin the process of developing a usable and practical test of MT ability. A two-pronged approach was taken to better understand (1) complex MT environments and (2) existing measures of MT. First, four MT environments were studied to begin to understand the cognitive operations they demand. A preliminary ontology of cognitive operations required by MT was developed and used to analyze the environments. The results of the analysis of MT environments were used to establish preliminary requirements for a predictive test of performance in those settings.

Second, a review of the literature was conducted to (1) identify current measures that could potentially be used to predict MT performance in real world settings, and (2) analyze those measures to determine the kinds of cognitive operations they measure. The literature was also thoroughly combed to garner a selection of existing laboratory measures for study. A selection of measures extracted from the literature were examined and evaluated to determine if they might form the basis of an MT ability test. The measures were analyzed using the same ontology of cognitive operations used in the analysis of MT environments. This permitted comparison of the cognitive processes tapped by measures and required by MT environments. The results of the analysis of existing MT measures was then used to select the best measure on which a test of MT ability might be based.

To begin the process of developing a usable and practical test of MT ability, current standards for educational and psychological tests were studied. Based on four phases of test development prescribed by the standards, a plan for development of an MT ability test was created. Following the plan, the initial phases of test development were completed. The purpose, scope, and framework for the test is described in this report and the test specifications currently supported by empirical research are given as well.

This report also describes the additional research necessary for further development of a test of MT ability. A set of studies has been designed to lay the requisite empirical groundwork for test development and to examine the construct and predictive validity of the resulting test. These studies are fully described in this report.

ORGANIZATION OF REPORT

The remaining chapters of this report describe the methods, results, products, and conclusions of the present research. One of the first tasks undertaken in this project was a thorough review of the literature related to MT. The purpose of the review was, as noted above, to garner, examine, and evaluate existing measures of MT. However, the review also provided the opportunity to relate the present research to the extensive knowledge base produced over the past 40 years on capacity limitations of human

information processing. Hence, the following section of this introduction provides an overview of literature related to MT. It reviews work accomplished in related areas such as working memory, personality, and information coordination.

The second chapter describes the formal technical objectives of the research. We then turn to the methods used to examine four MT environments and the results and conclusions of our analysis in Chapter Three. In Chapter Four, the methods, results, and conclusions of the analysis of existing measures are described. The conclusions of the two sets of analyses are then integrated in Chapter Five to describe gaps in the way MT is currently measured. Measurement needs revealed through analysis of MT environments and measurement capabilities provided by analysis of existing measures are then combined to begin the process of test design. In Chapter Six, we discuss current testing standards that have been used to guide initial development of a test of MT ability. The test specifications are given in this chapter. In the final chapter of the report a set of studies are described that may be used to further develop and validate the test.

OVERVIEW OF LITERATURE RELATED TO MULTI-TASKING

The capacity to perform multiple tasks within a limited time frame has been of theoretical and empirical study from nearly the onset of psychology as a science (Brookings & Damos, 1991). The objective of much of this research has been to understand the capacity limitations of human information processing. Researchers have sought to determine how individuals are able to control their mental operations in conditions of mental load and under time constraints. Experimentally, taxing the cognitive system through time pressure and information overload in a MT environment is a way to reveal the constraints of the system that may remain concealed when it runs unimpeded. By pushing the cognitive system beyond its limits, an opportunity is afforded to address some fundamental questions concerning the human cognitive architecture (Mayer & Kieras, 1997), such as the existence and functionality of a central processor (Pashler, 1994). From this long line of study, the limitations in MT performance have been linked to a number of theories about the cognitive system, including limited capacity structural bottlenecks (e.g., Broadbent, 1958; Deutch & Deutch, 1963; Pashler, 1994; Treisman, 1969), resource sharing (e.g., Kahneman, 1973; Navon & Gopher, 1979; Pashler, 1989; Wickens, 1984), and the flexible control of executive processes (e.g., De Jong, 1995; Meyer & Kieras, 1997). We have also learned a great deal about the kinds of tasks that can and can't be shared (Wickens, 1984).

Much of the early research related to MT primarily attempted to demonstrate the existence of attentional control while performing more than one task (Braun & Wickens, 1986). In a typical experiment individuals were asked to perform multiple tasks under two conditions: (1) a single task condition, in which one's full attention could be directed to the task, and (2) a dual-task condition, in which subject must divide attention between two tasks. The division of attention could be dictated by the experimenter to the research participant (e.g., 75% directed to Task A, 25% to Task B), or it could be totally left to the discretion of the research participant. Some dual task studies employed a correlational approach in which changes in the pairwise correlations between tasks done alone and in combination were expected to give rise to a latent

factor structure comprising single task abilities and a general time sharing ability (e.g., Stankov, 1983). Unfortunately, where positive evidence was provided, any general factor was usually less general across a variety of task combinations and instead was more specific to particular pairs of tasks (Brookings, 1990). Ackerman, Schneider, & Wickens (1984) provided a critical review of much of this research and admonished that the evidence for or against a timesharing ability was indeterminate based on lack of theoretical frameworks defining components of a timesharing ability, methodological flaws in experimental designs, and inappropriate statistical analyses.

The advent and ubiquity of new technologies in the 1980s was the impetus for the creation of new computer-controlled testing paradigms, and as such was viewed as an important advancement in extending the range of cognitive abilities measured in two important ways: (a) the modification and expansion of testing procedures of more familiar psychological functions and capacities, and (b) extending the testing of psychological functions and capacities to abilities not typically included in more conventional psychometric batteries (Hunt & Pellegrino, 1985). Such technological advances were predicted to benefit evaluation of individual differences in attention, particularly if the goal was to predict performance in complex situations entailing rapid decision-making.

Because of the previous difficulties in identifying a global time-sharing ability, subsequent research sought to discover the more basic information processing mechanisms for handling multiple sources of information that may lead to individual differences. Many of these research endeavors have been in the context of understanding the component operations and the capacity limitations of a working memory (WM) system. Another area of research has focused on MT situations in which multiple sources of information must be coordinated to perform a task or several tasks. The WM studies have focused on either (a) studying the operations of working memory in MT environments (i.e., loading a hypothetical WM component with a secondary task and observing the performance decrement) or (b) investigating the relationship of WM with higher forms of cognition that had MT as a characteristic of the task. The information coordination studies have attempted to find evidence of a general ability factor to integrate multiple sources of information. We will first discuss the WM research related to MT.

WORKING MEMORY CAPACITY

WM is important to performance in real world MT environments because it may constitute one architectural limit of the information processing system that constrains workload. A working memory system must prioritize and direct attention to multiple tasks to achieve accurate perception, situation awareness, and decision-making. Working memory as a psychological construct, has been related to MT by Joslyn and Hunt (1998) in their research on time-pressured decision making. The predictive power of a working memory updating task (Larson & Saccuzzo, 1989) on a simulated task of emergency dispatching and ATC was significant, accounting for 8% of the variability in the DISPATCHER task and nearly 15% of the variability in the ATC task. Compared to the ADM task, however, this measure of working memory had far less predictive power. Researchers, however, have developed many measures of WM and it is entirely possible that the particular task used in this study does not adequately represent the

aspects of WM that are important to MT. Other measures of MT than the one used by Joslyn and Hunt (1998) might better predict performance in an MT environment. For example, memory updating is one hypothesized component of WM that may be integral to most MT environments.

Baddeley's Working Memory Model. According to Baddeley's (1996, 2000; Baddeley & Logie, 1999) influential model, the construct of working memory refers to a limited capacity information processing system that is responsible for the simultaneous storage and processing of information during the performance of a variety of cognitive tasks. Such processing is said to be invoked through the interactions of two temporary storage components (the phonological loop and the visual-spatial sketchpad), plus a supervisor that oversees and controls the online processing of the entire system. This supervisor, known as the central executive (CE), operates to coordinate the products of the two slave systems and to integrate multiple sources of information. As such, it is hypothesized to be responsible for the rapid redeployment of mental resources in order to supervise complex cognitive processing. The vast majority of the early research on Baddeley's model focused on the two storage systems, which confirmed and extended the traditional notions of STM (e.g., Atkinson & Shiffrin, 1968). In contrast, the CE only recently has received substantive research. In its original conception, Baddeley modeled the CE on the supervisory attentional system (SAS) offered by Norman and Shallice (1980). The SAS serves as a dynamic and adaptable controller for resolving competition and promoting cooperation among cognitive processes. In essence, the SAS is said to be involved in any activity that requires strict attentional regulation. Baddeley (2002) now assumes that the CE component of his model is purely attentional in nature, attributing to the CE three primary functions dealing with the capacities to focus, divide, and switch attention. A number of studies presented below support the hypothesis of attentional control by relating the CE to other (higher) forms of cognition. Although subsequent research has taken a variety of approaches (e.g., experimental, cognitive modeling, neurophysiological) and demonstrated a variety of different viewpoints on how a CE might work in cognitive activity, all ascribe the CE with some function of attention for controlling mental operations.

Executive Functions. Engle and colleagues have developed an extensive program of research aimed at investigating the various phenomena and functions of working memory. The scope of this line of research ranges from understanding the relationship working memory capacity has with higher forms of cognition (e.g., fluid intelligence) to disseminating the critical functions of working memory executive functions. For instance, Engle, Tuholski, Laughlin, & Conway (1999) demonstrated through structural equation modeling that WM, as measured by complex span tasks (e.g., reading, operation, and counting span tasks) is functionally and statistically distinct from STM, as measured by simple forward and backward digit and word span tasks. After partialing out the common variance between STM and WM, the residual variance was highly correlated with fluid intelligence, as measured by Ravens Progressive Matrices and Cattell's Culture Fair Test. Engle et al. argued that this relationship between WM and fluid intelligence was predicated on the common demand for controlled attention. More precisely, an executive aspect of memory operates in controlling attention for the purpose of maintaining activation of goal-relevant information while inhibiting goal-

irrelevant information. They conclude that WM capacity might best be conceptualized as a system comprised of STM capacity and an executive component responsible for controlled attention.

Engle and colleagues have further supported their view of controlled attention as the key feature of WM by relating operation span (OSPAN) performance to tasks that have a substantial element of attention with minimal memory storage demands. More specifically, individuals with high WM span performed significantly better than low WM span individuals on several different attention-control tasks, including dichotic listening tasks (Conway, Cowan, & Bunting, 2001), antisaccade tasks (Kane, Bleckley, Conway, & Engle, 2001), and Stroop tasks (Kane & Engle, 2003). These attention-control tasks effectively require the inhibition of habitual processes as an element critical to performance.

Miyaki, Friedman, Emerson, Witzki, Howerter & Wager (2000) provided evidence for the separability of three select CE functions of working memory. These targeted CE functions were (a) mental set shifting (or attention switching), which entails the intentional disengagement of an irrelevant task set and the subsequent engagement of a relative task set, (b) information updating and monitoring, which involves actively manipulating (encoding, revising, replacing, tagging, sequencing) information, and (c) inhibition, which necessitates the deliberate suppression of prepotent responses. Through latent variable analysis Miyake et al. demonstrated that the three CE functions are statistically distinct. In addition, they examined the roles of these CE functions in more complex executive tasks, with mental set shifting most closely related to performance on the Wisconsin Card Sorting task, updating and monitoring most closely related to performance on both Random Number Generation and Operation Span tasks, and inhibition most closely related to performance on both Tower of Hanoi and Random Number Generation tasks.

Suss, Oberauer, Wittmann, Wilhelm, and Schulze (2002) formulated three working memory functions and explored these in the context of intelligence. The first working memory function pertained to the simultaneous storage and processing of information. Such a function was analogous to the global definition given by Baddeley (1996) for working memory, and was measured by Suss et al. with various span tasks. The second function was a supervision function, which might be best considered a set of functions falling under the rubric, "executive functions". These included monitoring and controlling the efficiency of mental operations, activating appropriate schemata, and inhibiting inappropriate schemata. Suss et al. operationalized this supervision function with verbal, numerical, and figurative switching tasks. Finally, Suss et al. proposed a coordination function that operates in integrating isolated pieces of information into new coherent structures. This ultimately requires simultaneous access to multiple, distinct pieces of information for the purpose of using them as elements in new relationships. This last function was measured using a number of memory updating tasks. Ultimately, the data revealed that the storage/processing function and the coordination function were statistically non-distinct. Using confirmatory factor analysis, a non-orthogonal two-factor structure for working memory was derived comprising storage/processing/coordination as one factor and supervision as another. Both of these reliably predicted global intelligence, with the storage/processing/coordination factor being most closely

aligned with reasoning abilities and the supervision factor being most closely aligned with speed of processing.

The research presented above is representative of the opinion that working memory executive functions primarily deal with the control of attention. Furthermore, the control of attention comes in a variety of forms including inhibiting the processing of information, monitoring and updating the contents of thought, shifting attention between tasks or information sources, and coordinating or integrating information across tasks or information sources. As detailed next, the control of attention as it pertains to MT has a long and storied past that has made significant advances in the past decade.

Executive Control and Task Switching. Historically, the study of a general ability to MT has its roots in the study of attention. As past research has shown, however, the concept of attention has continually eluded precise quantification and definition. Questioning the direction of research in the psychology of attention, Allport (1993) admonished that a unified theory of attention is little more than wishful thinking and instead one should be resigned to the idea that there are many different kinds of attention that serve a variety of cognitive processes. Focus should therefore be directed at characterizing the diversity of attention. One cognitive process that is arguably central to an ability to MT concerns the control of attention. Individuals in MT situations often must rapidly engage and disengage attention to multiple information inputs as the situation demands (Wickens, 1999). Early research has demonstrated that the ability to switch attention (a) has external validity with other complex, multicomponent tasks like aircraft piloting and bus driving (Gopher, 1982; Gopher & Kahneman, 1971; Kahneman, Ben-Ishai, & Lotan, 1973), (b) includes both general and modality specific characteristics (Lansman, Poltrock, & Hunt, 1983), (c) functions in the processing of externally and internally derived sources of information (Hunt, 1986), and (d) is implicated in working memory control processes (Carlson, Sullivan, & Wenger, 1993). Of particular note is the research by Gopher (1982) in which an attention switching measure was incorporated into an already established pilot selection battery that included other measures of attention. The findings indicated that those cadets who successfully graduated from flight training school consistently performed better on measures of attention, with the attention switching measure demonstrating the greatest difference. Furthermore, the attention switching measure significantly contributed to the prediction of flight school success, whereas other measures of attention did not. Because pilots need to efficiently attend to appropriate information and be able to rapidly redeploy attention to appropriate stimuli, the timing of events is critical, with both tardy and inefficient switching of attention to rapidly changing conditions leading to deficient performance.

Research interest in the control of attention waned somewhat through the late 1980s and early 1990s. However, recent research interest has been directed at addressing questions regarding the regulatory processes underlying supervisory control functions. At the heart of this is a renewed interest in the study of attention switching as an element of cognitive control (e.g., Baddeley, Chincotta, & Adlam, 2001; Carlson, Sullivan, & Wenger, 1993; De Jong, 1995; Emerson & Miyake, 2003; Meiran & Marciano, 2002; Meyer & Kieras, 1997; Pashler, Jolicoeur, Dell'Acqua, Crebolder, Goschke, De Jong,

Meiran, Ivry, Hazeltine, 2000; Sohn & Anderson, 2001). According to Gopher, Armony, and Greenspan (2000), cognitive control emphasizes the manner in which individuals configure/reconfigure tasks, choose among alternative subgoals, and monitor and adjust mental effort in order to optimize performance. As an element of cognitive control, attention switching is a cognitive activity that is utilized in everyday life as individuals routinely switch among tasks, trains of thoughts, and multiple information sources. When switching, more fundamental cognitive operations are invoked including selective attention, inhibition, and the temporal sequencing of mental operations.

Sohn and Anderson (2001) have made the argument that task switching requires both executive and automatic forms of control. That is, switching that is under executive control is endogenous in nature and refers to the intentional, goal-directed switching between mental activities. On the other hand, switching that is under automatic control is exogenous in nature and is driven by certain conditions (i.e., stimuli) in the environment. In MT situations this distinction is important because individuals in some MT environments may be subject to more or less endogenous and/or exogenous control. That is, as detailed above, some situations may be rife with interruptions from the environment that must be dealt with immediately (i.e., exogenous task switching), whereas other environments may allow more volitional choice of which of many tasks to do (i.e., endogenous task switching). Ostensibly, some of the task switching in MT environments is strategic in nature, whereas other forms of switching may be considered more reactive to a demanding environment. The former can be linked to the planning, prioritizing, decision-making, and prospective memory features of many MT environments (Burgess, 2000).

A number of accounts of executive control in working memory have been applied to experimental MT situations (i.e., dual task experimental paradigms) with an emphasis on strategic processing. For instance, De Jong (1995) recognized that a higher-order control structure may supervise MT performance. Central to this control structure are preparatory strategies that schedule the performance of multiple tasks, as well as regulate and arrange for the timely switch to, and subsequent processing of, other tasks. Accordingly, a central control mechanism critically functions in preparing for performing multiple tasks. Advanced preparation for either retrieving or implementing appropriate performance strategies facilitates more continuous forms of processing between multiple tasks. Thus, preparatory strategies are said to reduce or prevent any competition for limited capacity mental structures, as well as exploit opportunities for the temporal overlap in MT processing.

The strategic control of MT processing has been formalized in a production system simulation by Meyer and Kieras (1997), with the model accurately accounting for systematic individual differences in a number of MT performance situations. Their Executive-Process Interactive Control (EPIC) architecture is a computational framework that attempts to model MT processing with an interactive production system comprising perceptual, cognitive, and motor processes, as well as a set of executive processes regulating the interplay of the three. The executive processes schedule and control the operation of task-specific rules, monitor task progress, and shift task priorities. Such executive actions interact with the task-specific processes by placing appropriate information into working memory and/or by inducing anticipatory switching between

tasks. Executive processes regulate the progress of multiple tasks by monitoring the partial outputs deposited into working memory. According to the EPIC framework, various scheduling strategies must ultimately be invoked to manage the performance of multiple tasks. At one extreme is a lock-out algorithm that dictates strict sequencing of multiple tasks and allows no temporal overlap in multitask processing. Alternatively, an interleaved scheduling algorithm may suspend the concurrent processing of multiple tasks for short intervals and allow component task processing to proceed with varying degrees of temporal overlap. The adaptive use of these two scheduling strategies is dependent upon particular task combinations, individual preference, and degree of practice/experience.

INFORMATION COORDINATION

Another psychological construct potentially related to performance in MT performance that warrants further study is information coordination (IC) (Yee, Hunt, & Pellegrino, 1991). IC is a unique instance of the dual-task experimental paradigm in which two related component tasks must be concurrently processed, with the products of such component task processing integrated under time constraints. Successful performance in an IC task requires mental processing specific to each component source of information plus the real-time integration of the two component sources of information (i.e., coordination). The capacity of individuals to effect such an integration appears to be distinct from their abilities to process each component source of information. Because of the need to relate information from one component task to another, the IC situation requires more precise control of mental operations than does a standard dual task. A number of individual differences studies have consistently demonstrated the existence of an IC ability contributing to overall individual variability. Furthermore, a number of important issues concerning IC have been broached including simple practice and task complexity (Morrin, Law, & Pellegrino, 1994), extensive training (Law, Morrin, & Pellegrino, 1995).

A convergence of evidence for the role of some form of coordination in MT performance has come from experimental, neuropsychological, and developmental perspectives. For instance, Emerson, Miyake, and Rettinger (1999) importantly extended the Yee et al. (1991) research on information coordination by demonstrating that performance of multiple related (i.e., coordination) tasks was correlated with performance of multiple unrelated tasks (i.e., standard dual tasks). Emerson et al. also manipulated the degree of temporal overlap for the multiple tasks, finding that MT abilities were directly linked to the degree of temporal overlap in executing multiple tasks, a phenomenon also observed by Morrin (1996). Finally, Emerson et al. found that both related and unrelated MT performance correlated with a measure of attention switching. Emerson et al. concur with the conclusions of Morrin (1996) that MT situations critically invoke working memory executive abilities that operate in some information management capacity. Preeminent here is the ability to judiciously engage and disengage attention (i.e., attention switching) between competing sources of information.

Many instances of working memory research have made use of simple and complex memory span tasks. The former has been shown to be related more to the construct of STM whereas the latter has been associated with WM (Engle et al., 1999). In investigat-

ing the information processing properties of complex span tasks, Bayliss, Jarrold, Gunn, and Baddeley (2003) found that individual differences in complex span performance is attributable to both storage capacity and processing efficiency, plus an additional source of individual differences concerning the coordination of the two. That is, complex span performance requires the independent contributions of storage, processing, and executive coordination. Furthermore, executive coordination was related to adult fluid reasoning and reading and math skills in children.

Measuring activation levels in the cerebral cortex in a normal population (i.e., non-brain damaged), D'Esposito, Detre, Alsop, Shin, Atlas, & Grossman (1995) revealed that the coordination of multiple tasks requires the activation of additional brain areas (e.g., prefrontal cortex) that are not activated when tasks are performed in isolation. Frontal lobe patients with dysexecutive syndrome and patients suffering from dementia of the Alzheimer type (DAT) routinely demonstrate a substantial impairment in MT performance compared to performance on single tasks (e.g., Baddeley, Della Sala, Papagno, & Spinnler, 1997).

Finally, from a developmental perspective, Mayr, Kliegl, and Krampe (1996) have explored the role of coordinative processing as a determinant of lifespan developmental differences. Coordinative processing, in which information flows between interrelated processing components, was distinguished from sequential processing of simple, independent processing components. The former cognitive function required various aspects of task scheduling and task switching, as well as the timely reactivation and transformation of information across component processes. In their research, a developmental dissociation was found between basic processing efficiency (i.e., speed) and coordinative efficiency (i.e., working memory functions), with older adults significantly impaired in tasks requiring coordinative processing.

The Joslyn and Hunt (1998) studies included the IC task of Yee et al. as a predictor of their simulated ATC and dispatcher tasks with mixed results. That the IC task used by Joslyn and Hunt was only marginally related to the ATC (3% shared variability) and DISPATCHER (11% shared variability) tasks should not devalue its potential as a predictor. The apparent lack of relationship between IC and the complex simulations used by Joslyn and Hunt may have been due to the fact that the IC task used was severely time-constrained (less than 2 seconds). This may have artificially depressed the accuracy and thereby reduced performance variability. Other measures or IC tasks may show a stronger relationship to simulated real-world jobs that require MT. For example, Morrin (1996) has successfully used a composite score of accuracy per unit time to index individual differences in a different set of IC tasks and found that such a measure performed better in correlational analyses than either simple accuracy or response time. In sum, IC can be viewed as a capacity functioning in the control of attention during the time-critical management of multiple information sources. It appears to use the cognitive operation of real-time integration of two pieces of information plus operations specific to the processing of each task.

A compelling argument can be made that individual differences in MT performance may be strongly tied to WM executive functions. The literature suggests that WM uses several cognitive operations which would include manipulating and updating of infor-

mation, coordination among dual tasks, inhibition of selected information sources, monitoring task progress and partial outputs deposited in working memory, scheduling and control of task-specific rules, shifting task priorities, and shifting of attention. Though it is recognized that executive functions are not of one kind, the strategic control of attention as personified by attention switching may be the best predictor of performance across a range of MT situations. That Joslyn and Hunt (1998) found a relationship between a WM updating task and performance in a simulated MT environment, may just be the tip of the iceberg concerning the relationship between WM and MT. Such a relationship, in theory, may be more predictive than even the ADM task.

PERSONALITY FACTORS

Within the disciplines of industrial-organizational psychology, social psychology, and personality psychology, there are numerous investigations of individual differences in personality and job-related performance. By comparison, in cognitive psychology there have been relatively few studies looking at the relationship of performance in MT environments to dimensions of personality. Several of note focus on the impact of extraversion, Type A behavior patterns (TABP), impulsivity, and self-efficacy on MT performance. The research findings presented below suggest that non-cognitive dimensions of individual differences should be further explored to determine their predictive validity in MT jobs.

Extraversion. The personality dimension of extraversion has been linked to MT performance by looking to the physiological underpinnings of this trait. Arousal theories of extraversion (e.g., Brocke, Tasche, & Beauducel, 1997) depict the association between arousal and performance as an inverted-U relationship. As such, there are optimal levels of arousal necessary for optimal performance; too little or too much arousal results in suboptimal performance. According to Eysenck's (1997) view of personality, the dimension of extraversion also has a biological basis connected to physiological arousal. More specifically, introverts have higher baseline levels of cortical arousal and greater reactivity to environmental stimulation than do extraverts. Eysenck argues that extroverts may tend to compensate for their suboptimal levels of arousal by seeking greater stimulation from the environment in a variety of ways. In theory, baseline levels of arousal for introverts reside closer to optimal arousal levels compared to extraverts. Supporting this, extroverts have been shown to outperform introverts in tasks that substantially increase arousal levels, because that change moves extroverts into the optimal arousal-performance zone, whereas introverts are pushed to the downside of the inverted-U (Paisley & Magnan, 1988).

This arousal-cognition relationship receives additional support from cognitive neuroscientific studies investigating catecholamine (dopamine and norepinephrine) activity in the prefrontal cortex, an area implicated in studies of MT (Burgess, Veitch, de Lacy Costello, & Shallice, 2000), as well as the central executive component of working memory (D'Esposito & Postle, 2002; Rypma, Berger, & D'Esposito, 2002; Kane, 2002). More specifically, too much or too little catecholamine activity undermines attentional and working memory processing thought to be involved in MT situations. From all accounts, MT situations are likely to increase levels of arousal. For introverts, such a condition should produce an excessive (i.e., nonoptimal) amount of catecholamine

activity in the prefrontal cortex, thus impairing the introvert's effectiveness in MT. In contrast, MT should raise the levels of arousal in the prefrontal cortex to more optimal levels in extraverts, thus facilitating or improving their MT effectiveness.

The extraversion, arousal, performance relationship has been further confirmed in a MT setting by Lieberman and Rosenthal (2001), who hypothesized that extraverts should perform better in MT situations than introverts under two assumptions. First, MT is a skill that necessitates the efficiency in which working memory can control, inhibit, and invoke various competing goals. Second, MT situations are characterized as situations in which levels of arousal are elevated. From the rationale above, such situations may overstimulate catecholomine activity of the prefrontal cortex and thus subvert attentional and working memory efficiency requisite for MT performance. Lieberman and Rosenthal found that the performance by introverts in MT situations was impaired relative to that of extraverts. Further, they observed that extraversion was correlated with behavioral measures of central executive aspects of working memory but not associated with storage capacity.

Type A Behavior Pattern (TABP) Evidence has come from several sources supporting the idea that individuals demonstrating Type A behavior patterns (TABP) might be better suited for performing in certain MT situations where time pressure is an inherent quality of that environment. The TABP can be characterized by competitiveness, achievement striving, and time urgency (i.e., having the feeling of being under time pressure). Several studies have examined individual differences in TABP relative to MT performance. The research by Mathews and Brunson (1979) characterized TABP individuals as "hyperalert" in terms of appropriately directing their attention to task-relevant information while suppressing task-irrelevant information. More specifically, TABP individuals were more precise in controlling their attention in a MT situation, as well as better able to focus their attention when performing a single task while inhibiting distracting stimuli.

Subsequent work by De la Casa, Gordillo, Mejias, Rangel, and Romero (1998) looked at attentional strategies of TABP individuals in MT situations, focusing on how individuals prioritize their information processing. In a MT situation in which one task was designated as primary (i.e., standard dual task), TABP individuals demonstrated a greater intensity of focal attention to that task compared to Type B individuals. However, in an ambiguous MT situation, in which instructions were not given for one of the tasks, TABP individuals displayed an effective division of attention over the two tasks. De la Casa et al. concluded that TABP individuals exhibit better focus of attention directed at task relevant information when necessary (e.g., dual task), and distribute their attention better in ambiguous or ill-defined MT situations.

Finally, Ishizaka, Marshall, and Conte (2001) looked at the relationship of global TABP and the TABP subcomponents of time urgency (internally/self-imposed time constraints), achievement strivings (actively working hard to achieve goals), and polychronicity (the preference for working on more than one task at a time) with MT performance. Individuals had to perform in a MT situation comprised of three separate tasks (two visual and one auditory) in either unambiguous (i.e., full instructions given for prioritizing tasks) or ambiguous (i.e., incomplete instructions given for prioritizing

tasks) conditions. Global TABP measures were not related to MT performance, but time urgency and achievement strivings were. Additional research has revealed a relationship among time urgency, polychronicity, and achievement strivings. That is, individuals driven for success often take on more than one task at a time, performing with a sense of urgency in accomplishing their goals (Conte, Rizzuto, & Steiner, 1998).

Impulsivity. Schumacher, Seymour, Glass, Fencsik, Lauber, Kieras, & Meyer (2001) have shown that the degree of interference between two tasks in a MT situation can be modulated by instructions about task priorities and daring or cautious scheduling of tasks. Their research suggests that performance differences in MT conditions may, in part, depend on personality traits like impulsivity. The personality trait of impulsivity can be defined as the tendency to act with less forethought than most people of equal ability, with this lack of deliberation typically being seen as a negative quality in cognitive functioning (Dickman, 1990). However, there is some evidence to suggest that certain forms of impulsivity may actually facilitate cognition. For instance, in the context of very rapid decision making, high impulsive individuals have been observed to be reliably more accurate than low impulsive individuals (Dickman & Meyer, 1988). Dickman (1990) has more precisely refined impulsivity by fractionating the trait into statistically independent functional and dysfunctional forms. Functional impulsivity (FI) is said to facilitate performance in time-constrained conditions and is related to enthusiasm and productive risk-taking. In contrast, dysfunctional impulsivity (DI) is the tendency to engage in rapid, error-prone information processing because of an inability to slow down and more carefully process information when the situation allows for such an approach. Dickman (2000) went on to further demonstrate that impulsivity-related differences in cognitive performance reside in the ability to effectively allocate attentional resources.

Self Efficacy. Ackerman and Kanfer (1993) demonstrated that measures of self-efficacy reliably contributed to the prediction of ATC training performance independent of traditional cognitive ability measures. This finding held true for both laboratory and field settings. As such, Ackerman and Kanfer confirmed their hypothesis that self-reports of confidence would provide incremental validity to cognitive ability measures in predicting skill acquisition of a complex, attention-demanding task.

CHAPTER TWO: TECHNICAL RESEARCH OBJECTIVES

The **primary** technical objective of the present research was to **design a reliable and valid measure** of MT ability and the time it takes to achieve skilled MT performance. To support this technical objective, it was necessary to research complex real-world MT environments and existing measures of MT that might form the basis of a test. Hence, a **supporting technical objective** was to **examine the cognitive operations** required by jobs that require time-pressured MT such as military tactical decision-making and nursing. The product of the job analysis was identification of the cognitive operations performed by workers in a selected set of MT environments. A **second supporting technical objective** was to examine existing measures of MT to identify the cognitive operations they measure. The cognitive operations demanded by MT jobs were then compared to the cognitive operations measured by existing measures of MT to select an

existing measure on which the proposed test will be based. Figure 1 depicts the strategy used in the present research.

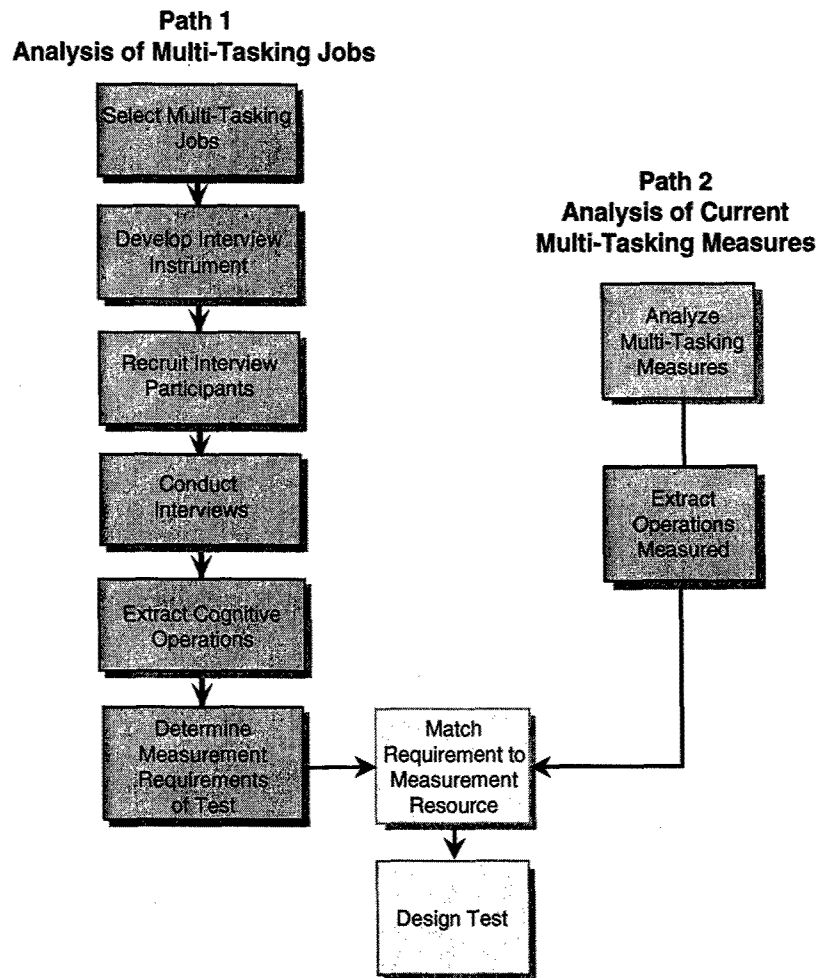


Figure 1. Research strategy to design test.

To meet the primary objective, initial phases of test development were completed, as prescribed by standards jointly developed by the American Educational Research Association (AERA), American Psychological Association (APA) and the National Council on Measurement and Education (NCME) (1999). The standards require that test development be grounded in empirical findings. Because very little research has been conducted in this area, it will be necessary to conduct additional research to meet the standards. Additional research will also be needed to establish the test's psychometric properties such as reliability, construct validity, and predictive validity. For this reason, a final **technical objective** was to **design a plan to validate the MT test**. Test development and validation studies will be completed in Phase II of the research.

CHAPTER THREE: INVESTIGATION OF MULTI-TASKING ENVIRONMENTS

Generally stated, the purpose of investigating MT environments was to gain knowledge about the kind of performance a test of MT ability should predict. This research constitutes an initial examination of the criterion performance the proposed test seeks to

predict. A better understanding of the similarities and differences among MT environments is imperative to development of a test that can predict performance in a wide variety of MT environments.

Several issues were important to this study. First, how similar and how variable are MT environments in terms of the kinds of cognitive requirements they place on individuals who work in them? Do they all require the capacity to remember lots of information, for example? Do they all require the interleaving of tasks, and hence the ability to use prospective memory? Is the ability to prioritize important to all MT environments? Which cognitive capabilities make someone good at these jobs?

Because very little is known about real-world MT jobs, it was also important to examine the similarities and differences in the external environment in which the jobs are performed. For example, what are the elemental tasks like in each job? How cognitively demanding are they? Is time pressure a factor in all MT settings? Is interruption present in all MT environments?

The general strategy was to analyze a sample of MT environments to identify their common characteristics and the cognitive operations that they appear to demand. Using the characteristics and the cognitive operations as bases for comparison, it was then possible to determine how the environments differ, and how they are similar. We reasoned that a general test of MT ability should incorporate the cognitive operations demanded by all, or at least most, MT environments. This strategy assumes that MT ability is largely cognitive in nature. Since most of the research in dual tasking, time-sharing ability, and task switching stems from cognitive psychology, the assumption that MT is largely a cognitive ability, or a set of cognitive abilities, is reasonable. However, it is also possible that performance differences among workers in MT environments may also be influenced by personality factors. While several studies have investigated the influence of personality variables such as TABP, confidence, and risk-taking (see overview of literature in this report), the relative contribution personality factors make to MT performance remains a question to be answered by future research.

In the present research, MT settings were studied by interviewing individuals who worked and had extensive experience in the jobs studied. A conclusive understanding of the cognitive requirements of MT environments would necessitate the use of additional research methods. For example, protocols might be taken while subject matter experts in a particular field worked on real or simulated tasks. Alternatively, experimental conditions might be devised that would conclusively demonstrate the need for certain cognitive processes but not others. Unfortunately, these methods were beyond the scope of the current project's resources. Moreover, because this study constitutes the first published research investigating cognitive requirements of MT jobs, protocol analysis and experimentation entail greater cost and risk than is appropriate at this stage of knowledge. Therefore, conducting interviews was judged the best method under the circumstances.

To focus the interviews, the critical incident technique was utilized. Participants were asked to describe incidents that they had experienced in which the MT demands on the job were particularly taxing. Hence, a set of jobs was examined at times when a

high demand was placed on MT resources. Because of the limitations of interview methods (e.g., they produce self report data subject to retrospective error and bias) the results of this analysis constitute only a preliminary view of the cognitive requirements of MT jobs. The results should be later tested and validated through other more rigorous methods.

SELECTION OF MT ENVIRONMENTS

The first step in this component of the research was to select a set of MT environments to study. However, what constitutes an MT environment? The literature does not include a consensus definition of MT, let alone a definition of the kind of environment in which it is demanded. In fact, only a few researchers have attempted to define MT (Burgess, 1998, 2000; Joslyn & Hunt, 1998). While it is quite easy to think of jobs that *probably* demand MT ability, it is not possible to a priori identify an MT job without careful examination guided by a clear definition of MT. For example, does the job of driving a racecar qualify as an MT job? The driver simultaneously receives lots of visual and auditory information from the environment and from radio communication to the pit crew. His or her progress is interrupted by other cars and simultaneous operation of several controls must be executed to be successful. However, any racecar driver will report that they are completely focused on *one* task while driving: driving. They experience a focused state of mind, not one distracted by numerous different tasks to be accomplished. Hence, the lack of a clear definition of MT and MT environments makes it difficult to determine whether racecar driving is an MT job or not.

DEFINING MT ENVIRONMENTS

Selection of a set of MT environments for study clearly requires definition of MT. As noted above, there is no consensus definition and few researchers have attempted to provide one. However, we found Burgess's (2000) approach to defining MT settings useful. Burgess et al. use the following characteristics to describe real-world multi-tasking situations. Note that Burgess is not attempting to describe the cognitive operations required by MT environments, only the environments themselves. (We have made comments in italics noting our own elaboration of the characteristic where appropriate.)

- Many tasks: Several tasks must be completed, which are discrete and different from one another. *Note: it is probably possible that some MT environments incorporate the same general tasks that must be repeated on different organizing units. For example, a nurse must deliver solid form medications (multiple instances of the same task) to several patients (the organizing unit in this example) using similar, if not identical, procedures. By definition, an MT environment must include multiple discrete tasks, but they need not be different from one another. However, we are quibbling because most real-world MT environments probably include different tasks.*
- Interleaving Required: Tasks must be interleaved because the environment does not permit the shedding or postponement of tasks so that another task can be performed and completed. *Note: interleaving of tasks is a strategy used by workers, not a characteristic of the environment. A better way of stating this characteristic is that the*

environment does not permit the shedding or postponement of tasks due to their importance or urgency.

- One task at a time: It is not possible to perform more than one task at a time because of physical or cognitive limitations. *Note: It may be possible to perform aspects of tasks concurrently in an MT situation. The inability to perform more than one task at a time may be due to the environment or due to the limitations of the human processing system. A better way of stating this characteristic is that all the necessary tasks to be completed cannot be simultaneously performed.*
- Delayed intention: The time for a switch or return to a task is not signaled directly by the situation. Hence, scheduling of tasks is up to the performer. *Note: MT environments are probably more complex than this characteristic affords. The time to switch to another task may be cued for some tasks in some MT environments. Again, delayed intention is not an environmental characteristic, but a response or strategy used by workers in the environment. A better way of expressing this characteristic is that the environment does not signal or cue scheduling of tasks.*
- Interruptions and unexpected outcomes: Unforeseen circumstances and interruptions of tasks will occur. The environment is uncertain in this way and is not under the control of the performer. *Note: interruptions are a specific form of a dynamic environment where information concerning tasks and the external world is constantly changing. Dynamism may be a common feature of MT environments.*
- Differing Task Characteristics: Tasks differ from one another in terms of priority, difficulty, and length of time. *Note: Some MT environments may include tasks that have the same level of priority, difficulty, and duration. That said, the vast majority of real-world MT environments probably include different tasks that vary substantially along these dimensions.*
- Self-determined targets: People must decide for themselves what constitutes adequate performance. *Note: this characteristic may be tantamount to the next one concerning the lack of feedback.*
- No immediate feedback: Errors or other indicators of performance may not be made available by the environment. *Note: while this characteristic may be true for some tasks, feedback may be provided for other tasks in many real-world MT environments. However, again we are quibbling because most MT environments probably include at least some tasks for which there is no feedback.*

Several characteristics might be considered for addition to this list of features. First, consider a time/task dimension in which tasks that take on the order of milliseconds to complete are placed at one end and tasks that take days or longer are placed on the other end. Real-world tasks that must be performed and interleaved within milliseconds are probably rare and may be beyond human processing capability. On the other end of the scale, tasks that require more than minutes, perhaps hours or even days probably would allow the shedding or postponement of tasks such that one or more tasks could be completed before another is attempted. Hence, most MT jobs probably require that several tasks must be performed over a period not exceeding a magnitude of minutes.

Second, most MT environments are probably time-limited simply because many tasks must be completed within a limited period of time. MT environments must have some kind of time limitation because otherwise there would be no reason to interleave or simultaneously perform multiple tasks. This characteristic seems necessary in defining MT environments to distinguish them from working settings in which tasks can be completed serially.

Third, the tasks in real-world MT environments probably vary in the amount of cognitive resources they demand. Some tasks may be performed automatically (e.g., steering a car) as they have been proceduralized, while others require focused attention (e.g., talking on a cell phone). Hence, tasks probably vary in terms of cognitive demands they place on the information processing system.

Finally, MT work environments require that workers be trained or educated. It is difficult to think of an MT job that could be performed by untrained individuals. Training and/or education are probably required.

Burgess (2000) provides an initial reasoned attempt to describe MT environments. The features he posits can be empirically tested by examining a sample of MT settings, which is the approach taken in the present research. To clarify and extend Burgess's original specification, Table 1 provides a revised list of eleven characteristics that define MT settings.

Table 1.
Defining Characteristics of MT Environments

#	Characteristic of MT Environment	Cognitive Operations Req'd by Environmental Characteristic	Rationale For Cognitive Operation
1	Multiple Discrete Tasks	Mental Set Switching STM storage STM rehearsal	PRP and task switching literature indicate that <i>mental set</i> must be changed when alternating between tasks. <i>STM storage</i> is necessary to remember completed tasks. STM storage typically requires <i>rehearsal</i> . May also require <i>planning</i> to organize multiple tasks in time and sequence
2	All the necessary tasks cannot be simultaneously performed	Mental Set Switching	Tasks must be sequenced in some way (serially, interleaved, or overlap) Requires task switching, hence <i>mental set switching</i> .
3	Tasks cannot be shed or significantly postponed because they are important or urgent	Prospective Memory	If tasks cannot be shed or postponed, and they vary in priority or duration, then they must be interleaved. If interleaving is used as a strategy, <i>prospective memory</i> is required to remember incomplete and future tasks.
4	Environment does not signal or cue task initiation	Prospective Memory	If tasks are interleaved and there is no cue to get back to or initiate a task, worker must use <i>prospective memory</i> .
5	The environment is dynamic and includes interruptions	Divided Attention Selective Attention WM Updating and Monitoring	Interruptions are a form of dynamic environment where information is coming in from a variety of sources. These would demand either <i>selective</i> or <i>divided attention</i> , or both. Dynamic environment where worker continuously receives information

			about tasks or their own performance of tasks requires constant <i>WM updating and monitoring</i> .
--	--	--	---

Table 1. (Continued)
Defining Characteristics of MT Environments

#	Characteristic of MT Environment	Cognitive Operations Req'd by Environmental Characteristic	Rationale For Cognitive Operation
6	Tasks differ in terms of priority, difficulty, and length of time	Prioritization Deductive Logic	This requires prioritization because some of the tasks are more urgent. <i>Prioritization</i> in turn uses <i>deductive logic</i> to establish priorities.
7	Feedback is not provided for some tasks	Classification Judgments LTM Retrieval	People must make judgments or <i>classifications</i> of adequacy of their performance. It would also require <i>LTM retrieval</i> on which to base judgments.
8	Most tasks are performed in the order of seconds to minutes	Prospective Memory	Must use <i>prospective memory</i> because tasks must be interleaved because they must be performed within minutes.
9	Environment is Time Pressured	Prospective Memory	Must use <i>prospective memory</i> because tasks must be interleaved because they must be performed quickly.
10	Tasks vary in the amount of cognitive resources they demand	Automatic Response Monitoring Prioritization	Some tasks are automatic, as execution has been proceduralized while others require focused attention. This means that task execution of <i>automatic responses must be monitored</i> and tasks that are demanding cannot be time shared and must be <i>prioritized</i>
11	Performance requires training or education	LTM Retrieval	<i>LTM</i> produced by training must be retrieved

It is important to distinguish characteristics of MT environments from the cognitive demands of MT, which Burgess and his colleagues also discuss (Burgess, 2000; Burgess, Veitch, de Lacy Costello, & Shallice, 2000). Environmental features of MT settings do not define MT as a psychological construct. For example, interruption is a feature of a setting, not a cognitive process. This is an important distinction because one might design a test to (1) simulate characteristics of MT environments, or (2) incorporate the cognitive processes required by those environments. Either strategy could be used to develop a predictive test of MT job performance. Unless there is an isomorphic relationship between environmental characteristics and cognitive operations, however, the two strategies might well produce very different kinds of tests that might differ in predictive power. Because it is unlikely that an isomorphic relationship exists, it makes the most sense to analyze jobs based on their cognitive operations rather than their environmental characteristics. By its very nature, MT ability is a cognitive construct. If the goal is to assess MT ability, the focus should be on cognition. Table 1 also provides a list of cognitive operations that are probably demanded by each of the environmental characteristics, as well as rationale describing the probably link. Later in this section of the report, we discuss an ontology of cognitive operations for MT, which is based on the rationale given in Table 1.

SELECTED MT ENVIRONMENTS

To begin the process of selecting MT environments for study, a list of candidate jobs was developed (See Table 2). Only those jobs of which the authors had personal knowledge, through previous research or personal experience, were included in the list. Hence, the list is not exhaustive, nor even representative, of all MT environments. Moreover, familiarity is no substitute for empirical study. Familiarity with the jobs afforded only initial positive judgments about the likelihood that they place workers in environments that demand MT. Hence, the list may well contain jobs that do not meet the characteristics noted above. While not exhaustive, the list seems to meet the purposes of the present research.

Table 2.
Candidate MT Jobs

Emergency Room Nurse
Emergency Medical Technician
Emergency Room Physician
Intensive Care Nurse
Floor Nurse
Waitress
Short Order Cook/Chef
Football or basketball coach
Television director of live sports broadcasts
Police officer
Fire fighting Captain
Stock broker
LCAC Craftmaster
LCAC Engineer
LCAC Navigator
Military Weather Reporter
Helicopter Pilot in NOE flight
Platoon leader
Company leader
Battalion leader
Brigade leader
Division leader
Navy anti-submarine warfare officer
Combat Information Center
Tactical Action Officer
Bridge Officer Aircraft Carrier
Air Officer (aircraft carrier)

Eight jobs performed in four very different work environments (given in bold font in Table 2) were selected from this list based on several criteria. First, we wanted to study a set of jobs that, on the surface, seemed to require different skills and knowledge. We

reasoned that study of how MT environments differ, as well as how they overlap, would broaden and enrich ultimate design of a test of MT ability. Hence, we wanted to maximize the variation among the MT environments we studied in this research. For similar reasons, we also wanted to study jobs that are performed by individuals who are at different levels of career development. We reasoned that this would help us to develop initial hypotheses about how cognitive requirements vary at different stages of one's career. For example, the cognitive requirements for a platoon leader, who is typically an Army Lieutenant, may be different from those demanded of a division leader (usually a General). A mix of military and civilian jobs was also desirable so as to maximize the applicability and commercial viability of the MT test we would develop. Ideally, we wanted to study jobs in several military services. We also wanted to study jobs that would most benefit from a selection test because (1) they experience a high turnover rate due to stress induced from MT, (2) they receive a large number of applicants, and (3) they would experience significant decreases in training and attrition costs. Finally, accessibility to populations (for both the current research and for future research) also played a role in deciding which MT environments to study.

Based on these criteria two military MT environments were selected: operation of the Navy's Landing Craft Air Cushion (LCAC) and Army combat unit command. Both the Craftmaster and the Navigator positions aboard the LCAC were investigated. Previous research (Stuster, 2001) had shown that the Craftmaster on the LCAC carried the highest workload. Hence, we initially believed that study of this position would be the most informative. However, initial interviews indicated the Navigator position might actually demand higher levels of MT. We thought that looking at the differences between the two positions within the same MT environment might be an interesting comparison. Moreover, initial interviews revealed that the Navy had experienced significant Navigator attrition (70%) during and after training because of the stress induced from MT. Hence, individuals who had performed either the Craftmaster or Navigator functions on the LCAC were interviewed. Three levels of Army combat command were also investigated: platoon, company, and division. This allowed us to generate hypotheses regarding changes in MT requirements at different stages of one's career, as previously discussed.

Two civilian MT environments were selected: restaurant food preparation/chef and nursing. Both of these civilian environments experience high turnover rates, and financial losses in training costs, due to burnout. Both are in industries that might benefit from a selection test that would identify individuals who are unlikely to respond positively to MT environments. Both apparently demand high levels of MT ability. Initial interviews of chefs indicated that MT demands vary depending on position in the kitchen, type of restaurant, etc. Similarly, initial interviews with nurses suggested that MT demands are different for intensive care nurses than for floor nurses. Hence, individuals were interviewed who had performed a variety of food preparation positions, or who had been either an ICU nurse or a floor nurse.

INTERVIEWS

PARTICIPANTS

Nine professionals who worked in four different MT environments participated in the interviews. Each of the participants was highly experienced and qualified in their own field. Three of the participants had extensive experience as the Craftmaster and/or Navigator aboard the Navy's LCAC. One had retired from the Navy and was currently working for a civilian contractor, serving as an instructor in LCAC training programs. The other two were on active service and were currently serving aboard an LCAC. Two had operational experience as they had served aboard a LCAC during Desert Storm. The three had between 8 and 13 years experience. Another two participants were retired Army officers who had served in combat leadership positions at the platoon, company, or division echelon. One had retired from the Army as a Four-Star General and the other retired as a Lieutenant Colonel. Two participants were nurses who had worked in intensive care units and/or medical/surgery departments of hospitals. One had 2 and the other had 14 years of experience. The final two participants had worked as professional chefs, one for 2 years and one for 10 years.

The nine professionals were recruited using an informal network of contacts developed by the authors through previous research studies in the areas of LCAC operation, military leadership, nursing and medicine, and food preparation. Civilian participants were paid a small honorarium of \$75 for their time. Active duty service personnel volunteered their time.

INTERVIEW QUESTIONS AND TECHNIQUE

A standard set of questions was designed to probe the cognitive requirements of any work environment, regardless of the particular field of work or job content. The questions were designed for use in the context of a critical incident of MT that the participant had experienced as part of his or her work. After describing the incident, the interviewer asked a series of questions pertaining to six different topics related to the cognitive requirements of the job including issues of memory, task prioritization, decision-making, knowledge and experience, the work environment, and relationships among the components tasks.

Questions about memory requirements probed the need for rehearsal, the existence of external memory aids, and the kind and amount of information stored in memory (prospective or retrospective). By definition, the reported incidents involved multiple tasks. Hence, the second set of questions involved how those tasks were prioritized, whether the participant had control over prioritization, and whether prioritization was important to performance. Issues concerning the kinds of decisions that were made, how those decisions were made (speeded pattern recognition based or more lengthy problem solving and deliberation), and the basis for decision-making were covered in a third set of questions. The importance of an extensive knowledge base and years of experience to performance were probed in a fourth set of questions. Fifth, questions about the characteristics of the MT environment were asked such as the presence of interruption, the ability to control interruptions, the ability to shed tasks, and the need to interleave tasks. Finally a number of questions about the tasks themselves were asked

including their duration, number, complexity, difficulty, similarities and differences, relationship to one another, and the presence or lack of feedback.

INTERVIEW PROCEDURES

After initially agreeing to participate in the interviews, participants were sent an informed consent form that fully described the purposes and procedures of the study. After returning a signed copy of the consent form, a phone meeting was scheduled. During the interview, the interviewer first explained the purpose of the study and the general strategy to be used in the interview. The purpose and procedure of the critical incident method was described. Participants were told that it is a method that tends to increase the recall of detailed information. They were then asked about their experience and qualifications for their job.

The critical incident was elicited by asking participants to remember a particular time when they were required to perform many tasks concurrently. They were told that we were not necessarily soliciting an incident in which an error or accident had occurred, nor were we interested in an incident in which unusually high levels of performance were demonstrated. Participants were asked to simply recall and describe an incident in their job in which they had many things to do at once.

After they had completed their description of the incident, participants were asked the series of questions described previously. Clarifying questions were also asked when necessary. Several participants recounted more than one incident as time allowed. All interviews were tape-recorded.

AN ONTOLOGY OF COGNITIVE OPERATIONS USED IN MT

The MT environments and jobs studied in this research are, on the surface, very different. Some have physical components and require psychomotor and visual-perceptual skills (e.g., chef, LCAC operation) while others do not (e.g., division echelon battle command). Some are military while others are civilian. They all require vastly different knowledge bases, different experience, and different training. One would not expect a chef to successfully perform the LCAC Navigation job, nor vice-versa. Yet the descriptions provided later in this report will convince most readers that they all are MT settings. In this sense, the four jobs studied in this research are similar.

Establishing the similarities and differences among MT environments requires a common basis for comparison, however. A basis that is grounded in cognition is desirable because we assume that MT is fundamentally a cognitive ability. What is needed is a coherent, consistent, and well-organized set of cognitive operations that are described at a level of description that could be used to distinguish MT environments from settings that do not demand MT. Stated in a different way, what is needed is an *ontology*, or a statement of the existence, of a set of cognitive operations that might be demanded by MT. An ontology like this would serve as a preliminary definition and testable model of MT ability.

The utility of this set of MT cognitive operations would not be limited to distinguishing environments. With additional development, it might also be used to determine whether, or how heavily, a particular job requires MT. Additional research could

determine if all the operations are equally important to MT. Some operations might contribute more to making a particular job one that involved multi-tasking. In other words, a scale could be developed so that one could rank jobs in terms of how demanding they are of MT. A coherent set of cognitive operations could also be used to evaluate measures that purportedly assess MT ability. They could be used as basis of comparing tests and laboratory tasks of MT.

Unfortunately, no one has yet developed a coherent model of cognitive operations used by the human information processing system, let alone a set that describes MT ability. The literature does have well developed and extensively studied cognitive architectures that, arguably, currently frame psychology's understanding of cognition (e.g., EPIC, Meyer & Kieras, 1997; ACT-R, J. R. Anderson, 1993). One might turn to these architectures to identify a potential set of cognitive operations for MT. This strategy makes sense because computational models based on the cognitive architectures have been developed for applied and laboratory MT tasks. For example, an ACT-R model has been successfully developed for the MT job of an Anti-submarine warfare coordinator aboard an AEGIS ship (Anderson, Bothell, Douglass, Haimson, Sohn, 2002; <http://act-r.psy.cmu.edu/workshops/workshop-2002/talks/>). Using EPIC as an example, one might create the following list of cognitive operations: proceduralize, receive input from physical sensors, send the outcome of sensory analysis to working memory, test conditions and executing actions (a production-rule interpreter), select and send symbolic responses to the vocal and manual motor processors, prepare and initiate movements, update the contents of working memory by adding and deleting goals, steps and notes, and program the motor processors. The problem with this list is that it doesn't provide a way to *distinguish* one environment from another. *All* environments, whether they require MT or not, demand these cognitive operations. When is proceduralization or any of the other operations on this list *not* used in real-world tasks? The level of description for cognitive operations taken from cognitive architectures is too low to distinguish environments or potential measures of MT. It may be possible to derive a set of cognitive operations based on the task assumptions incorporated by an existing computational model of MT performance. However, it is not clear whether that derivation would produce any better ontology than analyzing the MT environments themselves. It would not be clear, for example, which operations were necessary to MT and which were not.

The present research took an alternative empirical, bottom-up, approach to specifying cognitive operations demanded by MT settings. First, the characteristics of MT environments were specified based on clarification and revision of previous research (Burgess, 2000; Burgess, Veitch, de Lacy Costello, & Shallice, 2000), as given in Table 1. Second, the cognitive operations those characteristics must require were specified. Burgess and his colleagues have also taken this approach, and once again we found their work to be useful (Burgess, 2000; Burgess, Veitch, de Lacy Costello, & Shallice, 2000). For example, Burgess et al. (2000) identify prospective memory, planning, and retrospective memory as important cognitive operations demanded by MT situations. The ideas presented in this report represent an extension of Burgess et al. (2000).

Table 1 shows the cognitive operations that should be required in MT settings based on the defining features of MT environments developed for this research. Twelve cognitive operations that we posit are necessary to the performance in MT environments are listed. As noted previously, several of these operations have been discussed by other researchers. For example, Burgess (2000) identifies retrospective memory and intentionality as ability dimensions that predict performance in MT tasks. Intentionality is the ability to follow one's plan and the task rules, which is similar to prospective memory. It is interesting that Burgess (2000) also propose that *planning* is an important ability to successful MT performance, which we do not give in Table 1. By "planning" he means the ability to form a plan, which is a statement of how one intends to complete a set of tasks. A plan would specify the sequence and duration of each task to be completed. Hence, a plan is something one creates *before* the remaining required tasks are begun. The list of MT environmental characteristics, however, does not necessarily require the ability to form a plan. One must prioritize among tasks, but one could do that without benefit of a plan in the sense that Burgess uses it. One might deduce that planning is required if one made certain assumptions about the MT environment that we do not make. If one assumes that it is possible, for example, to accurately predict at least some events in real-world MT settings, and if time is available before the other tasks may be initiated, then it might make sense to develop a plan that would specify the sequence and duration of each task, for example.

The list of twelve cognitive operations cannot be considered to be complete or exhaustive. We propose this as a preliminary model of MT ability that should be tested in future research. This preliminary model was used, however, to compare the four different MT environments studied in this research. The results of this analysis are described in the next section of this report.

INTERVIEW RESULTS

Individuals who work in four MT environments were interviewed. Based on their responses to the interview questions and the incidences they reported, descriptions of the four environments were derived. The eleven environmental characteristics and the related twelve cognitive operations noted in Table 1 were then used to analyze each environment. The remainder of this section of the report is organized according to the four environments studied in this research.

DESCRIPTION OF LCAC ENVIRONMENT

The LCAC is a vehicle used by the Navy in performing amphibious assaults. This hovercraft operates at high speeds from launch points over the horizon and can deliver equipment and personnel to the world's beaches without the need of hydrographic surveys of boat lanes. The LCAC is contained within the well-deck of a mother ship while deployed until it is needed for a mission. The crew aboard the LCAC is composed of five specialists who work together as a team to operate their high performance craft. The Craftmaster (operator), Engineer, and Navigator occupy the upper level or flight deck of the starboard cabin. The Loadmaster and Deck Mechanic are in the port cabin. The deck of an LCAC is an extremely dangerous place because of the propellers, turbine

engines, and high surface speeds. For this reason, all passengers and crew must remain inside the relatively small port and starboard cabins while under way.

The Craftmaster is responsible for operating the LCAC and providing leadership to the crew. The Engineer maintains and monitors the performance of all onboard equipment, and equipment related logs and inventories. The Engineer is also responsible for directing the crew's response to fires and other emergencies and serves to assist the Craftmaster. The Navigator plots courses and maintains and monitors the navigation equipment. The Navigator also is the crewmember responsible for maintaining the personnel, training, and event logs. The Loadmaster is responsible for developing load plans, securing deck cargo, and monitoring the status of cargo while under way. The Loadmaster also serves as a port side lookout. The Deck Engineer works closely with the Engineer during start up and shut down.

All five LCAC crew wear headsets and microphones to remain in constant communication with each other from pre mission inspection through post mission shut down. Intra crew communication is an essential part of the LCAC work and it constitutes a primary source of MT for the crew. The crew also receives extensive radio communication from the mother ship, other ships in the vicinity of operation, and other LCAC.

We interviewed three individuals who had extensive experience on the LCAC (1 Craftmaster and 2 Navigators). The MT demands vary with crew position; hence, we consider the Craftmaster and Navigator positions separately here.

Craftmaster. The Craftmaster is responsible for operating the LCAC and providing leadership to the crew. His primary responsibility is to control the craft's velocity and direction between an amphibious assault ship and the assigned destination ashore. He takes heading and speed input from the Navigator to guide operation of the craft. While operating the craft controls, he continually scans the external environment through the starboard cabin window, which is typically wet from sea spray. He must also visually scan his instrumentation to assess the craft's current status. He receives communications from the other members of the crew aboard the craft and from sources external to the LCAC, such as the mother ship. He frequently receives direction updates from the Navigator, visual reports from the Loadmaster, and craft status reports from the Engineer. He receives radio communications internal to the craft on one side of his headphones. External communications are delivered in the other ear of his headphones. Operation of the craft is frequently interrupted, but cannot be postponed, by communications. If the mother ship, for example, attempts to communicate on the radio to the Craftmaster during a particularly difficult maneuver, e.g., quick avoidance of an obstacle, the Craftmaster may postpone responding to the call until after the maneuver is completed. In this way, operation takes priority over other tasks.

Most of his tasks are maneuver tasks, but he is also responsible to respond to emergencies. The most critical tasks the Craftmaster performs include well deck entry with support ship at anchor or underway, operate craft in a variety of weather conditions, traverse slopes such as sand dunes, translate land to water during surf conditions, tow another craft, and respond to and direct crew response to craft fire (Hunt, Linnville, Stuster, Schneider, & Braun, 1993). The most important abilities a Craftmaster must

possess are visual-motor skills. He must have excellent reaction time, depth perception, and spatial orientation. He must also have excellent night vision and near vision. Leadership skills such as teamwork, assertiveness, and comprehension or information given orally, are also paramount for a Craftmaster.

Navigator. The Navigator serves as a filter for all pieces of information pertaining to navigating the LCAC for the Craftmaster. His primary responsibility is to communicate navigational information such as heading and speed to the Craftmaster so as to adhere, as much as possible, to the planned route. He prepares the planned route ahead of time based on the mission. However, the plan is only a starting place because missions typically do not go by plan. When unforeseen events occur, the Navigator must recalculate heading and direction to ensure that the craft arrives at its appropriate destination on time.

He has many sources of information that he uses to accomplish this basic task. He constantly monitors an on-surface radar screen looking for potential obstacles that might interfere with the craft's progression toward its destination. He also visually skims the horizon as well as the cargo deck through the cabin window, looking for potential obstacles and using the visual information to establish situation awareness. The Navigator also has other instruments to which he must attend. For example, current heading input and velocity are shown on the displays to which he must continuously attend. The Navigator also has a GPS system that he uses to correlate the location of the craft with paper charts that he has. Although the GPS system greatly facilitates awareness of spatial location, there is always the possibility that the GPS will malfunction or go out completely. Hence, the Navigator is always checking the GPS information by dead reckoning navigational means. The paper chart has a great deal of information on it concerning the specifics of the mission, which also assists the Navigator in maintaining situational awareness. He uses the chart for updating purposes as well by writing current position on the chart as well as heading and speed information. In fact, updating of location, speed, heading are a continuous process for the Navigator.

While constantly scanning his environment, he simultaneously receives communications from other ships, other crafts, the beach and other crewmembers. The mother ship may warn him of potential obstacles that they pick up on their surface radar, for example. He must monitor communications directed to him and other crewmembers on 5 different radios. At any one point in time, he may need to speak to several different people. External communications are given in one ear while internal communications come over the other ear.

While performing his navigational and communication tasks, he is also responsible for recording information into the craft logs. The logging responsibility is no small task, as all events of any significance must be recorded as well as information about craft speed and heading during the mission. The log is used as one of the primary records of the mission. Hence, it is a very important tool that is used in researching mishaps when they happen. The Navigators we spoke to often found themselves logging information simultaneously with talking to another crew member of craft on the radio. The Navigator is also responsible for mission planning. Before a mission, the Navigator briefs the entire crew about the mission.

The Navigator prioritizes his various tasks, first addressing those that will most interfere with the mission. Borrowing a priority mnemonic from aviation, the Navigator aviates, navigates, and then communicates, in that order. Hence, the first priority for the Navigator is collision avoidance. If he is assured that the craft is safe from obstacles with which the craft might collide, he then focuses on making sure the LCAC is reaching planned intermediate way points at the right location and time. Internal (among the crew) and external (with other craft) communications are given a third priority.

The typical LCAC mission is to deliver a payload to a beach, which may require several trips between the mother ship and the beach destination. Timing and location is critical. Arrival at the target destination is severely constrained. The LCAC missions require that the craft arrive no later than 3 minutes after the planned arrival time, and not earlier by any amount. The craft must also be positioned within 500 feet of the planned destination. To meet these strict goals, the Navigator must constantly reassess the craft's position relative to the planned position and planned time, which often involves calculating distances, velocities, and headings. The Navigator is concerned with all obstacles, but classifies them as either (1) critical, which requires a change of direction or maneuvering around, or (2) of passing interest, which does not require any change to the navigational plan. When the craft leaves the mother ship, the Navigator must make note of the ship's location, velocity, and heading at the time of departure because he will have to find it after dropping the payload on the beach when the LCAC makes the return trip back.

The most cognitively demanding task that must be performed by the Navigator while underway is caused by any sort of maneuvering off the planned route. When an obstacle requires that the craft take an unplanned turn, for example, the Navigator must re-compute the whole navigational picture. The Navigators that we interviewed told us that the mental number crunching required to recompute required heading and speed was the hardest part of the job. If, for example, an obstacle required that you alter course, it may open the distance to the beach. Because timing is critical, the Navigator then must figure a way to compensate for the additional distance that must be covered. He might increase speed to 50 knots from the planned 35 knots to reach the next control point at the scheduled time, or he might figure that he needs to increase speed to 38 knots throughout the entire mission. Either way, he must perform calculations on the fly to give the Craftmaster the appropriate heading and speed that will accomplish the mission. The mental calculations are sufficiently difficult that they should not be interrupted by other tasks. Navigators do not simply punch in numbers in a computer to derive an answer to their navigational needs. Most of these calculations are done mentally. If they are, the Navigator may have to start all over again, but if so, when he does start, the situation will be even more different than before because the craft will have been moving in some direction and the clock will have been ticking. It is possible to be pushed into even greater error, which can be catastrophic to the mission, when mental calculation is interrupted. Under heavy cognitive demand, the Navigator may postpone making entries into the deck log, and may also ignore radio communications.

The Navigator's biggest asset is preparation. The more information about distances between control points, timing, and speed that he can compute before the mission and have written on his charts, the less figuring he has to do should the Craftmaster ask him a question. For example, at any point in time the Craftmaster may ask the distance to the next waypoint. If the Navigator has computed and noted distances between waypoints in preparing for the mission, he may then simply answer by checking the information he's recorded on his paper chart, thereby avoiding the need for recomputing the information on the fly. This is a strategy he uses to reduce working memory load. He uses this technique and other memory aids as much as possible, but much simply has to be remembered.

LCAC ENVIRONMENTAL CHARACTERISTICS

The environments in which the LCAC Craftmaster and Navigator work fit the eleven characteristics noted in Table 1 for MT settings. They both involve multiple tasks that cannot be simultaneously performed or shed. Their tasks differ in priority, difficulty, and length of time and in terms of the cognitive resources required. They both determine when they will perform each task as none of their tasks are signaled or cued by the environment. At the end of a mission, they both must review their decisions to evaluate their performance because the environment does not provide feedback for each decision they make, although it does tell them whether they've met the mission requirements in broad terms. They both face a very dynamic environment that includes interruptions. Most of their tasks are performed within a magnitude of seconds. However, one difference is that some of the visual-motor tasks the Craftmaster must perform are probably executed in less than a second. They are under time pressure because they must not be later than 3 minutes to their destination on the beach. Both are extensively trained.

Although sitting right next to each other on the LCAC, their MT environments do differ somewhat. This difference is not reflected by the binary system we used here, however. The biggest difference is the sheer number of different tasks that the Navigator must perform exceeds those of the Craftmaster. The type of skills required by the two jobs also differ in that the Craftmaster's task tap visual-motor skills and the navigator's involve higher level cognitive skills such as problem solving and calculation. The individuals we interviewed reported that the Navigator's job involves MT to a much greater degree because his tasks are more different from each other than the Craftmaster's, there are many more of them, and they demand greater cognitive resources.

LCAC COGNITIVE OPERATIONS

The cognitive operations required by the Craftmaster and the Navigator positions also differ somewhat. In determining whether each job required each cognitive operation, we sought independent evidence based on descriptions of each job given in the interviews and responses to questions directly addressing the cognitive operations. In this section of the report we list each of the twelve cognitive operations specified in Table 1 and provide examples of how they are required, or not, by each job. We first discuss the Craftmaster position.

Craftmaster Cognitive Operations

1. Retrospective Memory (STM)

- Must remember heading and speed directions he receives from Navigator
- Must remember information about obstacles he receives from Navigator and other external sources
- Must remember communications from the Engineer about craft status

2. Retrospective Memory (LTM)

- Continuously uses knowledge of craft capabilities
- Continuously uses knowledge of craft operation

3. Prospective Memory

- Does not use prospective memory extensively because most of his tasks are cued by the Navigator, other crewmembers, or the environment. Also, his tasks primarily serve one overriding mission, which is to operate the LCAC, which entails updating his situation awareness of craft status. Hence, the number of tasks he has, and has to remember, is small. The only evidence of prospective memory was that the Craftmaster might occasionally decline to respond to a communication, which he had to remember to get back to at a later point in time.

4. Monitoring Output

- Uses automatic visual-motor response to guide craft, which he must consciously monitor to ensure accuracy, especially in conditions where craft guidance is difficult, such as in entry into the well deck.

5. Working Memory Updating

- Continuously updates situational awareness of status of craft relative to planned mission

6. Mental Set Switching

- Must switch attention to different tasks, e.g., communications to scanning, to making changes in velocity or heading

7. Classification

- Does not use classification extensively except to determine whether performance has been adequate or not. For example, he has either hit an obstacle or not, which is a relatively trivial classification.

8. Rehearsal

- Did not report using rehearsal to remember STM items. If he needs the information again, he asks the Navigator.

9. *Selective Attention*

- Must at times attend to only one communication, for example, the Navigator's directions

10. *Divided Attention*

- Must continuously monitor five radio channels
- Must operate craft controls while performing visual scans of environment
- Must perform visual scan of instruments while operating craft controls

11. *Prioritizing*

- Places operation of the craft as the highest priority over other tasks
- May postpone communications until after attention demanding maneuvers are completed

12. *Deductive Logic*

- Did not report using deductive logic.

Navigator Cognitive Operations

It is important to note that in addition to the following list of cognitive operations, the Navigator is also responsible for planning the mission, which is a cognitive operation noted by Burgess (2000), but not included in our ontology. The Navigator plans in advance of the mission anticipating information the Craftmaster will need and develops a navigational plan.

1. *Retrospective memory (STM)*

- Must remember heading, speed, and location of mother ship when last departed

2. *Retrospective memory (LTM)*

- Draws on knowledge of craft capabilities
- Continuously uses computational knowledge and skills

3. *Prospective memory*

- Must remember to return to interrupted logging task
- Must remember to periodically scan cargo deck, instruments, horizon
- Must remember to return to interrupted radio communications

4. *Monitoring Output*

- Did not report using automatic responses.

5. *Working Memory Updating*

- Continuously updates situation awareness, which includes information about heading, speed, next control point, location, etc.
- Continuously updates understanding of craft location relative to other objects such as ships, other LCAC, obstacles, beach, etc.

6. *Mental Set Switching*

--Must switch among very different types of tasks such as logging entries, calculating distances, headings, or times, and visually scanning instrumentation

7. *Classification*

--Must classify potential obstacles as ones that require maneuvering vs. ones that are only of passing interest

8. *Rehearsal for memory storage*

--May rehearse STM stores such as heading, speed, and location of mother ship when last departed so as to figure the current location of mother ship and return

9. *Selective Attention*

--Must attend only to calculation task when being performed
--Must attend to obstacles when present

10. *Divided Attention*

--Must log and talk on radio at same time
--Must scan horizon and talk on radio at same time

11. *Prioritizing*

--Must prioritize his many tasks, and typically place priority on those that involve maneuvering of the craft

12. *Deductive Logic*

--Uses deductive logic to figure locations of other potential obstacles given their headings, speed, and original location
--Uses deductive logic to refigure heading and speed of craft so as to keep to original navigational plan

DESCRIPTION OF ARMY COMBAT COMMAND ENVIRONMENT

We interviewed two individuals who had important experience in Army unit command during combat operations. One had commanded at every level from company through division. He retired as a four-star general as commander of Forces Command (FORSCOM). The other, who had retired from the Army as a Lieutenant Colonel, provided descriptions of incidents from his combat experiences as a reconnaissance platoon leader. The environmental characteristics of platoon, company, and division command during combat and combat training situations differ significantly, as do the MT demands they place on commanders. Hence, we discuss them separately.

Division Command. The division commander is responsible for providing leadership and tactical direction for the division, which is composed of three maneuver brigades,

an aviation brigade, a brigade-sized division support command, and a brigade-sized division artillery unit. In addition to these six major subordinate units is a mélange of engineer, signal, air defense, and military intelligence battalions. During combat, he has competing responsibilities. On the one hand, he must avoid careless risk. He represents the best opportunity for a unity of effort throughout his heterogeneous unit. Recklessly exposing himself to physical danger imperils that all-important unity of effort. On the other hand, he must go where on the battlefield he judges that he is needed, or where he must go to observe first hand what cannot be easily communicated to him. For instance, to understand the progress and present circumstances of a maneuver brigade, there is no substitute for him to look into the eyes of the brigade commander and adjudge not only the facts he is being presented, but the attitude and spirit of the brigade commander. Also pressuring the division commander forward is the knowledge that the soldiers – who are facing danger every day – must see him forward.

When he is present in his main combat headquarters, he is at the one location where all information relevant to division operations is designed to congregate. Here, he may obtain the best overall picture of present operations and the best thinking of each of his functional experts. In reality, the main headquarters is a cacophony of noise and competing priorities. Each of the functional experts – including artillery, air support, logistics, aviation, maneuver, engineers, et al. – are resolving difficult tactical issues, many of which could have critical implications on the overall operation. As he walks into this headquarters, the commanding general is greeted by a collection of subordinates, each of whom believes that he/she has a critical report that demands his immediate attention. These well-intentioned, well-qualified experts contribute to the MT environment that the CG must navigate.

The division commander must be an expert on his unit and its integration into the battlefield at hand. He must be able to distinguish between the immediately critical and the potentially critical reports from his multifarious experts. Next he must understand how to “buy time.” Some tasks *are* more important than other tasks, although all seem urgent. While all may be important, some are critical. Once important tasks pass over the threshold to become critical tasks, one must allocate time to deal with each of them. A decision maker must explore several options with time critical decisions before him. One, “Do I have to decide now? Two, if I must decide now, can I make a partial decision that will “buy time” so that I can move to the next decision?” The process of setting priorities takes into consideration both the relative criticality of each decision and how vital time is to each. Finally, the division commander must have the ability to focus on the problem he has fenced time for. Our interview subject remembers actually declaring to his staff, “Give me time to think!” He would divorce himself from the immediate on-goings around him to focus, analyze, and decide.

Company and Platoon Command. Company commanders are responsible for leading and directing three platoons. Although each has a small staff that helps with logistics and ancillary combat support skills, those staff members are virtually unavailable during active engagement with an enemy. The one “staff officer” the company commander can count on in the midst of battle is the company fire support officer – a field artillery lieutenant. They are responsible for making tactical decisions at the company level necessary to meet their mission. They are closer to the line of

engagement than is the division commander and are often under direct fire. Their responsibilities include the direction of the platoons during movement to battle and their employment in battle.

Platoon leaders have parallel responsibilities with the company commander, but one echelon down. The primary units of their command are the three squads that comprise the platoon. They have no staff, but generally do have an artillery sergeant who can directly request artillery and mortar support. They are the front line and are in mortal danger. The platoon leader's responsibilities include the direction of the squads during movement and the employment of the squads during battle.

During combat operations, both company and platoon leaders receive multiple communications from both higher and lower units. They must communicate and coordinate with other companies and platoons to direct force on the enemy and to severely limit fratricide. Constant radio communication is a feature of all combat situations. The purposes of the communications may be status reports, reports of enemy sightings, requests for resources, and questions about further action. The combat environment at all levels is extremely dynamic. It is characterized by multiple and simultaneous events, problems, and situations. The battle rarely goes precisely as planned, and a company or platoon commander is typically faced with soldiers to rescue, unanticipated enemy location or resources, reports of land mines, and other situations or events to deal with. These events are unpredictable and typically occur concurrently. Hence, the company and platoon leader is usually faced with multiple situations to resolve.

ARMY COMBAT ENVIRONMENTAL CHARACTERISTICS

The environments in which Army division, company, and platoon leaders work are described well by the eleven characteristics noted in Table 1 for MT settings. Multiple tasks must be completed that cannot be simultaneously performed or shed. For example, leaders at each echelon must simultaneously monitor and make decisions about multiple ongoing situations. Combat leaders face tasks that differ in priority, difficulty, and length of time and in terms of the cognitive resources required. Some decisions, for example, are immediate and nearly automatic (e.g., return of enemy fire). While others (e.g., tactical responses to ongoing situations) engage problem solving skills that may take minutes to even hours to complete. Although the dynamism and seriousness of the situation means that nearly all of a combat leader's tasks are urgent, a competent commander learns which tasks have a higher priority. Some tasks in the environment are cued, particularly at the platoon level. For example, enemy fire is an environmental cue that may require immediate response (e.g., return of fire). However, the environment does not cue all tasks so that combat leaders decide when they will perform many of their responsibilities.

There are many paths and plans by which any mission may be accomplished. Hence, even if the goals of a mission have been met, a leader must evaluate his performance to determine if it was met in the best way possible. The environment may provide immediate feedback for some actions the leader takes. However, the environment typically provides only vague feedback that must be interpreted and

evaluated. Combat command at all echelons is a highly dynamic environment that includes multiple interruptions. The division commander has greater control over interruptions than do commanders at lower echelons because the consequences of his actions are played out over a longer duration. Therefore, interruptions may be postponed or delegated to his staff. At each echelon studied, most tasks take seconds to minutes to perform. However, the division commander may take several hours to perform some tasks that are cognitively demanding (e.g., problem solving tasks). Time pressure is an inherent component of combat as timing of task execution may determine the outcome of battle. The Army provides extensive training for commanders of all echelons. In particular, the division commander is provided with many years of training and experience, which he draws upon extensively to perform his duties.

ARMY COMBAT COMMAND COGNITIVE OPERATIONS

The cognitive operations required by the division, company, and platoon leader positions also differ somewhat. Here we list each of the twelve cognitive operations specified in Table 1 separately for division, company and platoon leaders.

Division Command Cognitive Operations

1. *Retrospective Memory (STM)*
 - Must remember previous he has orders given to staff
2. *Retrospective Memory (LTM)*
 - Draws upon extensive knowledge base concerning strategy and tactics
 - Draws upon extensive knowledge base concerning enemy capabilities
 - Draws upon knowledge about enemy to make predictions about enemy intentions
3. *Prospective Memory*
 - Division commander's aide serves prospective memory role. In this sense the division commander does not keep in mind the complete set of multiple demands placed on him. He uses aide for that function to allow him to focus fully on each task.
4. *Monitoring Output*
 - Did not report need to monitor output and results of his automatic responses.
5. *Working Memory Updating*
 - Continuously updates understanding of multiple situations as they develop
 - Monitors progress toward mission and updates understanding of that progress
6. *Mental Set Switching*
 - Switches between leadership tasks and decision-making tasks

- Switches among decisions about tactics and decisions about strategy
- Switches between receiving updates from various units, to making decisions and delivering orders, to cognitively demanding problem solving concerning tactics or strategies

7. *Classification*

- Uses knowledge of common tactics and maneuvers to classify enemy actions

8. *Rehearsal*

- Did not report using rehearsal to store information in STM

9. *Selective Attention*

- Sometimes orders staff and aides to provide uninterrupted time for cognitively demanding tasks
- May focus on one situation at the expense of others if it is given a high priority

10. *Divided Attention*

- Receives reports from multiple sources
- Monitors multiple situations as they unfold

11. *Prioritizing*

- Must use experience to prioritize the many decisions he must make, as all are urgent
- Must prioritize among multiple situations to monitor and make decisions about

12. *Deductive Logic*

- Uses extensive deductive logic in strategy and tactics

Company Command Cognitive Operations

1. *Retrospective Memory (STM)*

- Must remember previous orders he has orders given to staff and unit
- Must remember CDR intent
- Must remember mission statement
- Must remember placement, battle plans, missions, etc. of other companies in battalion

2. *Retrospective Memory (LTM)*

- Draws upon stored knowledge about weapons capabilities, enemy characteristics, tactics, etc.
- Draws upon knowledge obtained in leadership training provided by Army

3. *Prospective Memory*

- Must remember to return to postponed communications, e.g., from his battalion command
- Must remember to monitor ongoing platoon situations
- Must remember to allocate resources or make decisions about ongoing situations

4. *Monitoring Output*

- Must sometimes inhibit automatic responses, e.g., if receiving enemy fire, the automatic response is to return fire, which may not be the best tactic at the time

5. *Working Memory Updating*

- Must continuously update his situation awareness of battlefield
- Receives periodic updates on each of his platoon's situations, which he uses to update his understanding of the battlefield
- Receives periodic updates from higher up, battalion, which he uses to update his understanding of the battlefield

6. *Mental Set Switching*

- Must switch from communicating with platoon to receiving radio messages from battalion
- Must switch from making resource allocation decision to calling battalion requesting artillery to using problem solving skills to decide best tactics
- Must switch from leadership tasks that promote morale and unity in unit to decision-making tasks

7. *Classification*

- Must decide if fire is enemy or friendly
- Must use spot reports to determine kind of enemy unit he is facing

8. *STM Rehearsal*

- Did not report using rehearsal to keep information in STM

9. *Selective Attention*

- May choose to not attend to some communications, e.g., from battalion, during intense combat or while performing other tasks of higher priority

10. *Divided Attention*

- Must divide attention between multiple radio communication from platoons, battalion, or other companies
- Must divide attention between reports of events within multiple ongoing situations

11. *Prioritizing*

- Must decide which of many decisions he must make has the highest priority, e.g., request artillery vs. update situation awareness of forward platoon
- Must decide which of multiple leadership tasks vs. decisions he must make has the highest priority

12. *Deductive Logic*

- Uses deductive logic when making tactical decisions

Platoon Leader Cognitive Operations

1. *Retrospective Memory (STM)*

- Must remember previous orders given to soldiers
- Must remember orders received from Company
- Must remember location of other friendly units

2. *Retrospective Memory (LTM)*

- Draws upon stored knowledge about weapons capabilities, tactics, etc.

3. *Prospective Memory*

- Must remember to periodically update company on platoon's situation
- Must remember to monitor other platoons' situations
- Must remember to get back to postponed requests or communications from soldiers

4. *Monitoring Output*

- Must sometimes inhibit automatic responses, e.g., if receiving enemy fire, the automatic response is to return fire, which he may have been ordered to avoid

5. *Working Memory Updating*

- Must continuously update his situation awareness of his soldiers' situations
- Must update his understanding of other platoon's situations

6. *Mental Set Switching*

- Must switch between executing tactics to communicating with soldiers
- Must switch between making tactical decisions to performing leadership tasks to encourage morale

7. *Classification*

- Must decide if other unit he sees is enemy or friendly
- Must decide if air attack is enemy or friendly

8. *Rehearsal*

--Did not report using rehearsal to store information in STM

9. *Selective Attention*

--Must postpone communication with company commander during intense fire

10. *Divided Attention*

--Must attend to multiple simultaneous events that occur during battle

--Must communicate with his soldiers while executing tactics

11. *Prioritizing*

--Must prioritize tasks such as communication to battalion reporting his situation, commands to soldiers, executing tactics, etc.

12. *Deductive Logic*

--Uses deductive logic in tactical decisions

NURSING ENVIRONMENT

The nurses we interviewed spoke about their experiences working in intensive care units (ICU) and in oncology or medical/surgical floors of a hospital. Each of the participating nurses had worked in both environments and was able to compare them.

Floor Nursing. The oncology or medical/surgical departments of hospitals care for individuals who have cancer or who have other medical problems requiring surgery, respectively. Patients may be very sick in either department or they may be well on their way to health. In the hospitals in which our participants worked, the floor nurses were typically responsible for six or more patients. However, this number varies among hospitals in the United States.

Several factors make floor nursing an MT environment. First, the nurse may have to interleave several different kinds of procedures/tasks that must be performed for each patient. For example, he/she may have to set up an IV drip or an infusion pump to control delivery of fluid medication, check vitals, deliver orally delivered medication, respond to patient and family requests, teach patients and family members how to care for the patient during and after their hospital stay, perform a variety of medical procedures, or call the attending physician, to name just a few. During a visit to a patient the nurse may also perform a physical assessment by listening to heart and lungs, checking physical appearance, assessing alertness and orientation, and performing a musculoskeletal assessment. Charting much of this information is a requirement of their job, which takes a considerable proportion of their time. Because there is a limited amount of time with many responsibilities, nurses often do not complete a task before they start another one. Floor nurses are also responsible for educating patients and family members. For example, if a patient must continue treatment after their hospital stay, the floor nurse is the person responsible for teaching

the patient and family members how to do so. They must also answer family member's questions about the treatments, patient status, etc.

Second, the tasks associated with each patient must be interleaved among all the patients for whom the nurse is responsible. The nurses we interviewed told us that the sickest patients receive the highest priority and are seen first. The nurse must weigh and prioritize the needs of the patients. Some tasks require immediate response (e.g., a patient is not breathing) while others can be delayed (e.g., a patient has vomited and an aide can clean up and the nurse can check the patient later). Patients who are less sick will likely receive less attention when time is limited. Our participants reported that they tend to "scratch off" patients who are less sick, meaning that they are given less priority in the nurses' working memory.

Third, the nurse is frequently interrupted by events during a typical shift, which requires that he/she delay the current task to attend to one that has a higher priority. At any time during their shift, it is quite common to be interrupted by another nurse with a problem (e.g., a patient is vomiting or is in pain and the other nurse needs assistance), a patient (e.g., a patient is pressing the nurse call button), equipment alarm (e.g., the IV pump has completed its cycle and is alarming), or a family member who may have questions.

The nurses work in shifts that vary in duration, but usually last 8 to 12 hours. When a floor nurse arrives for his/her shift, the first task is to review the cases and recent events. Within the first half hour of their shift they receive updates from the previous shifts' nurse. According to the nurses we interviewed, they first determine who the sickest patients are, which enables them to prioritize their tasks. Nurses are responsible for delivering medications, which are typically given on a schedule, perhaps on the even hours. After reviewing the cases and getting updated on events, the nurse typically begins the process of delivering medications, for which there is usually a two-hour window in which they must be delivered. They obtain all the medications they need and then begin to administer them, first to the sickest patients and working down the priority list. It's not unusual to run over the two-hour window, which has the result of backing up all the other tasks the nurse must complete. The nurse is often in a situation in which he/she must engage in other tasks as well as continuing to run the medications in an attempt to get it all done.

Intensive Care Unit Nursing. Much of what was described for floor nurses is also true for the ICU environment. One difference is that the patients in the ICU are critically ill and require constant attention. For this reason, ICU nurses are responsible for only one or two patients. If something is going wrong with a patient, or if there are very few patients as would be true in a small rural hospital, there may be even more than one attending nurse. The critical nature of the ICU patient's illness requires additional medical procedures, which increases the workload compared to floor nursing. However, the patient load is lower and there is more restriction on family visits. Hence, the teaching load is reduced compared to floor nursing. An ICU patient is monitored with a greater number of medical devices, which reduces the amount of information that the nurse must monitor and keep in working memory. Vitals are continuously monitored and displayed by the monitoring devices. However, the severity of the

illness results in a greater number of problems and the requirement of urgency in task completion. During one of these urgent situations, the demand for MT may increase to a level that is more than one nurse can handle. To stabilize the patient, several nurses may be interleaving all of the tasks that the situation demands.

Per patient, there are more tasks that an ICU nurse performs than a floor nurse. For example, the ICU nurse may need to clear pressure lines, print out EKG strips, check the fluid medications that are hanging, double check medications against orders, monitor the patient's appearance, observe how the patient's vital signs change as a function the nurses' presence or actions, attempt to calm the patient so as not to increase the vital signs, check the equipment, set alarm limits, call the pharmacy and order drips, check the ventilator tube, suction the patient, check secretions and reactions, check drains, note their odor, color, amount, empty drains, change dressings, get supplies, perform neurological check, and chart much of the information they retrieve.

A typical shift, which would be 12 hours, would start with a general run down of whole unit. An ICU nurse might be assigned one patient if that patient has extensive needs due to the acuity of his/her illness. If a nurse is assigned to two patients, one of the patients typically has fewer needs. The ICU will first talk to previous shift's nurse and get a report on the condition of the patient. Then the nurse would perform a complete assessment on each patient for which he/she is responsible. Assessment in the ICU can take a considerable amount of time. It typically involves measures of neurological, cardiological, urine, skin integrity and bowel function, among others. After the assessment, medications may be delivered and laboratory tests may be taken. Family members may visit and the attending physician may call for updates. Care for the patient is also interleaved into these activities, such as bathing. The nurse may have to transport the patient to another location in the hospital for testing or medical interventions. Every two hours the ICU nurse must complete a full assessment.

Treatment for the patient is tailored for the health problem the patient is facing. Hospitals follow a treatment care plan devised for each kind of health problem, which consists of a set of goals for the patient. For example, the treatment care plan for a cardiac patient might include hemodynamic stability, good oxygen saturation, increase daily living activity without a corresponding increase on cardiac workload, and free of pain. The goals for any particular health problem are available in printed form, but are well learned by ICU nurses. Hence, ICU nurses focus their work on improving the state of their patients as indicated by the goals given in the care plan.

The tasks that ICU nurses perform range from very complex and delicate to routine. Some of the tasks require physical skill developed through practice and experience. Others heavily tap reasoning abilities.

NURSING ENVIRONMENTAL CHARACTERISTICS

The environments in which floor and ICU nurses work fit the eleven characteristics noted in Table 1 for MT settings. They demand nurses engage in multiple tasks that vary and cannot be shed or postponed. Nurses are sometimes cued by the environment to perform a task (e.g. an alarm goes off on a piece of equipment), but mostly they determine when they will perform each task. They are also responsible for determining

if they have met their goals that they themselves set for each patient. They must evaluate their own performance at the end of day, feedback is not given for each task. Nursing is a very dynamic environment in which new information is constantly being presented, including interruptions such as patient calls. Most of their tasks take minutes to perform, with none over an hour. They are under time pressure because they have too many tasks to perform in the time given. Nurses receive extensive education.

Both floor nurses and ICU nurses must multi-task in their jobs. However, floor nurses have more patients, while ICU nurses have fewer patients but more and different tasks for those patients.

NURSING COGNITIVE OPERATIONS

Because floor and ICU nursing require very similar kinds of cognitive operations, we consider them together here. Below each cognitive operation that we cite we provide example tasks that demand that particular operation. This should not be considered an exhaustive list of tasks, only ones reported by our participants that clearly indicate requirement of an operation. It is worth noting that both floor and ICU nurses make plans for their patients at the beginning of their shift. The ICU nurse plans a strategy with goals of improving the patient's vital signs over the duration of the shift.

1. Retrospective Memory (STM)

- Must remember medications delivered
- Must remember procedures administered
- Must remember each patient's case and recent events

2. Retrospective Memory (LTM)

- Must remember the procedures involved in performing each task, e.g., programming an infusion pump to deliver a volume of fluid medication at a particular rate as prescribed
- Draws extensively on knowledge (LTM) of physiology, effects of medication, disease, etc. This is particularly true of ICU nurse.
- Must integrate multiple sources of information to form a coherent understanding of patient's condition

3. Prospective Memory

- Must remember all the tasks to be performed on a particular patient without external memory cues
- Must remember to attend to the needs of many patients (floor nurse)
- Must remember to return to uncompleted tasks
- Must keep in mind all the non patient related tasks that must be accomplished before shift is over

4. *Monitoring Output of automatic Responses*

- Must monitor programming of infusion pump rate and volume and recheck medicine with physician prescription because over similarity among prescriptions and medications

5. *Working Memory Updating*

- Continuously monitors and assesses patient condition and updates condition in memory
- Continuously updates priorities as patient condition changes

6. *Mental Set Switching*

- Must switch set for each different task. For example, nurse must switch set between programming an infusion pump to checking patient vitals, to responding to an alarm or nurse call button

7. *Classification*

- Must use several attributes of the patient (e.g., blood pressure, color, respiration, temperature, etc.) to determine status. Overall patient status is classified in terms of degree of seriousness depending on these attributes

8. *Rehearsal for Memory Storage*

- May remind himself/herself of remaining tasks

9. *Selective Attention*

- May inhibit attention to lower priority tasks, e.g., phone ringing, device alarming, tasks that aide can do
- May ask others who are trying to communicate with them to wait while he/she is attending to someone else or to another task

10. *Divided Attention*

- Must continuously monitors patient condition while performing another task (e.g., setting up IV) she/he
- Must divide attention between measures of patient's condition
- May turn off alarm while tending to another task

11. *Prioritizing*

- Must prioritize tasks, and then place highest on the list those tasks that must be completed to facilitate the health of those patients that are the sickest

12. *Deductive Logic*

- Must apply deductive logic to assess patient status from attributes

RESTAURANT KITCHEN ENVIRONMENT

According to the chefs we interviewed for this study, the degree of MT that they perform depends on the area of the kitchen one is working. It also depends on the restaurant and time of day. A chef has no control over how fast the orders come in. At peak times, a busy restaurant has a very busy kitchen. In particular, the sauté station of a kitchen is extraordinarily busy.

One chef we interviewed stated that she was responsible for 12 different burners on the stove at the sauté station, which amounts to 12 different things cooking at one time. Concurrently, she was responsible for turning around to plate food as it comes up and is ready to go out, while continuing to cook other things. The chef may also be, as was one we interviewed, in control of two ovens. At peak times the ovens will have food items in them cooking and set to go off at different times, anywhere from 5 to 15 minutes. The chef must remember all that is cooking in their head, although they may have visual reminders of the items just by looking at the burners or the oven. The chef must know what to put on the fire or oven and what has to come out first. All of the tasks must be coordinated to produce a plate that is ready to be served, and all of the tasks must be coordinated with the rest of the kitchen. For example, the sauté station must coordinate with the cold food station.

Food preparation places considerable demand on memory. The chef must remember recipes for the many dishes available on the menu. Chefs typically do not use recipe books or lists. When the menu changes, the chef must memorize a whole new set of recipes and items on the new menu. Within the first few days of a menu change, performance is hampered by the need to inhibit memory for dishes on the previous menu and the weak memory for the new items.

There is also an organizer, called the "Expeditor" whose job it is to remind everyone in the kitchen staff what is needed when. The Expeditor controls the flow of the food coming out of the kitchen to be served by the waiters and waitresses. The chef cannot simply work at his/her own pace and put it up when it is ready. His or her product has to coordinate with the rest of the kitchen, which is organized by the Expeditor based on tables and orders. To give the reader a better understanding of the intensity of this MT environment, what follows is a brief excerpt from our interview with one chef.

They [the Expeditor] will talk back and forth accordingly to each area...you know they will say that the broiler person has 3 minutes, so Jennifer I need that up in 3 minutes... so you have feedback coming at you from everywhere....you've got tickets coming out on the line that come up in your window that you get, the Expeditor is talking to you, you are communicating with everyone else on your line to let them know where you are... and in all of that I am trying to plate things on one hand, I'm cooking things on 12 burners behind me, I've got 2 ovens right next to me

The following excerpt reveals how timing is everything in food preparation.

The tickets come in and they start off slowly. Say peak time is 7 o'clock. About 6:30 tickets kind of stroll in once every 5 minutes maybe, which is plenty of time. You are still not physically pumped up, you are ready, you have spent, however, many hours getting your station ready and things start coming in slowly and you start getting into the flow

of it. You work your orders one at a time at that point. Then at 7 or 7:15 people start coming in for their reservations and then all of a sudden you go from getting one ticket every 5 minutes to getting 10 tickets coming in every 2 minutes... and you may end up with anywhere from 30 to 60 or 70 covers all at once within a 20 minute window

Finally, the following excerpt speaks to the memory requirements of a chef.

Your brain is in 6 different places at once...just think about it... you have 12 burners on high heat with different things on each of those burners ... and you can have fish on one burner, which can be cooked in 2 minutes ... and you can have a steak on another burner that takes 15 minutes ... so your brain is constantly racing through those 12 burners ... you don't have a timer set ... in the middle of the rush there are no timers there is no one telling you ... in your head you are aware of all 12 of those places.

RESTAURANT KITCHEN ENVIRONMENTAL CHARACTERISTICS

A restaurant kitchen fits the eleven characteristics noted in Table 1 for MT settings. Chefs engage in multiple tasks such as plating the food and cooking on multiple burners that vary and cannot be shed or postponed. Chefs respond to a great deal cueing by the environment to perform a task (e.g., oven timer goes off) but mostly they determine when they will perform each task. Chefs do receive considerable feedback from the environment as well. For example, it's obvious when they've burned food. However, they have to evaluate their own performance for many other tasks for which feedback is not provided. Did the food taste good, for example? Was the decision to put the eggs on right after the potatoes the right decision? A restaurant kitchen is a very dynamic environment in which new tickets coming up, the expeditor is giving instructions, and other chefs are communicating new information. Interruptions are constant. Most of their tasks are performed on the order of seconds to minutes. They are under time pressure because they must get food out to the customers and they have many burners going at once. Chefs may be trained on-the-job or may attend schools. In either case, education specific to food preparation is required.

CHEF COGNITIVE OPERATIONS

Below we provide our assessment of the cognitive operations a chef must employ to meet the demands of a busy kitchen. Note that we have not included selective attention in this list because our chef participants indicated that focusing on one task (or one communication) at the expense of another was detrimental to performance in this environment. Those chefs who attempt to complete whole tasks tend to slow the kitchen down to a crawl. All communications to the chef are important in a kitchen and should not be inhibited or even postponed. In contrast, divided attention is extremely important to the job in that the chef must attend to as many sources of information that are possible. Also note that deductive logic did not appear to be used by chefs except to prioritize the timing of food preparation.

It is again worth noting that planning was not used by the chefs we interviewed. While chefs spend time organizing their station to prepare for an evening of work, they cannot control what is ordered and when it is ordered. For this reason, they do not prepare a plan for an evening's work before the orders starting coming in. They make

sure they have enough ingredients to prepare the foods that are on the menu. However, they do not develop a step-by-step action plan that they attempt to adhere to as their work ensues. With regard to planning, chef work is unlike the LCAC Navigator's work, which involves preparation of a detailed plan that prescribes a series of actions in sequence and coordinated in time. Nurses also prepare a plan for each patient as the start of their shift. Hence, "sticking to the plan" is not a concept that is relevant to chefs, although it is very relevant to LCAC Navigators and nurses. Chefs react to each new ticket without the benefit of planning.

1. *Retrospective Memory (STM)*
 - Must remember which orders he/she has fired (put on burners)
 - Must remember which cold food or orders he/she has plated
2. *Retrospective Memory (LTM)*
 - Must remember recipes
 - Must remember menu, which can change frequently
3. *Prospective Memory*
 - Must remember when to pull items off fire or out of oven
 - Must remember to plate items organized around a particular order or ticket
 - Must coordinate timing of future tasks with others in kitchen
4. *Monitoring Output of Automatic Responses*
 - Must inhibit the production of similar, but different foods.
 - Must inhibit production of foods that were on a previous menu when the menu changes
5. *Working Memory Updating*
 - Chef must continuously monitor and update progress of each item cooking
 - Chef must monitor and update progress of each ticket
6. *Mental Set Switching*
 - Must switch set for each different task, e.g., plating to monitoring food to firing food
7. *Classification*
 - Must monitor various attributes of cooking food including color, smell, consistency, and duration on fire to classify it as "done" or not
8. *Rehearsal for Memory Storage*
 - May talk to oneself aloud as a reminder of foods currently cooking, tasks to be performed, and tasks accomplished

9. *Selective Attention*

- Must be aware of all information about the food at all times. Chefs cannot afford to use selective attention. They cannot selectively attend to one dish and not another, nor can they ignore communications from other kitchen workers

10. *Divided Attention*

- Must divide attention between the numerous burners and ovens
- Must divide attention between what is cooking and communication in kitchen
- Must divide attention between what needs to be plated and what is cooking

11. *Prioritizing*

- Must coordinate timing of tasks to ensure that food is prepared in a synchronous manner with other kitchen stations and within a particular order. This requires prioritizing certain tasks before others

12. *Deductive Logic*

- Did not report using deductive logic.

MT ENVIRONMENT ANALYSIS CONCLUSIONS

ENVIRONMENTAL CHARACTERISTICS

As shown in Table 3, each of the four MT environments studied in this research appears to possess the eleven characteristics of MT settings originally specified by Burgess and further elaborated in this report. Our analysis confirms the idea that MT environments require workers to perform many discrete tasks, and they cannot perform all of them simultaneously. In most of the jobs, the tasks cannot be shed or postponed. A notable exception to this was that the Division commander might postpone certain tasks for several hours while engaging in another task that is particularly cognitively demanding. The relatively longer time duration of his tasks and, the fact that the consequences of his actions cannot be seen for an extended period of time, probably accounts for his capability to postpone tasks for a relatively lengthy time.

Table 3.
Summary of Environmental Characteristics of MT Environments
(√ characteristic found in MT environment)

#	MT Environment Characteristics	LCAC Operator	LCAC Navigator	Platoon Leader	Company Leader	Division Leader	IC Nursing	Floor Nursing	Chef
1	Multiple Discrete Tasks	√	√	√	√	√	√	√	√
2	All the necessary tasks cannot be simultaneously performed	√	√	√	√	√	√	√	√
3	Tasks cannot be shed or significantly postponed because they are important or	√	√	√	√	Sometimes postpones tasks for hours	√	√	√

#	MT Environment Characteristics	LCAC Operator	LCAC Navigator	Platoon Leader	Company Leader	Division Leader	IC Nursing	Floor Nursing	Chef
	urgent								
4	Environment does not signal or cue task initiation	Obstacles and terrain conditions are cues to a large degree	√	√	√	Cdr. has aide to cue tasks	√	√	√
5	The environment is dynamic and includes simultaneous interruptions	√	√	√	√	√	√	√	√
6	Tasks differ in terms of priority, difficulty, and length of time	√	√	√	√	√	√	√	√
7	Feedback is not provided for some tasks	√	√	√	√	√	√	√	√
8	Most tasks are performed in the order of seconds to minutes	Visual motor may be under seconds	√	√	√	May take longer than an hour for cognitively demanding tasks	√	√	√
9	Environment is Time Pressured	√	√	√	√	√	√	√	√
10	Tasks vary in the amount of cognitive resources they demand	√	√	√	√	√	√	√	√
11	Performance requires training or education	√	√	√	√	√	√	√	√

In each of the MT jobs we analyzed, the environment provided cues to initiate some tasks. In fact, many MT environments are full of indicators to initiate tasks. For example, communications such as radio transmissions in LCAC and combat environments are cues to listen and respond to the messenger. However, the MT environments we studied did not cue *every* task, which means that nurses, chefs, combat leaders, and LCAC crewmembers must determine when to initiate at least some of their tasks. Two notable differences should be discussed here. First, the LCAC operator tasks are cued primarily by the environment and the Navigator. The environment of obstacles and terrain conditions send immediate cues to respond by changing course, slowing down, etc. Second, the Division Commander actually has an aide to cue him to initiate certain tasks. Hence, in these two jobs environmental cues are present to a greater degree.

Dynamism was a critical component of each of the four environments. Environmental dynamism is critical because (1) it requires workers to continuously update their memory and understanding of the situation, and (2) it means that the worker must decide how to attend to multiple sources of information that the environment provides. Each of the four environments force the worker to deal with simultaneous presentation of an array of visual and auditory stimuli, which means that

workers must control attentional resources. He/she must allocate selective attention to those stimulus sources that are deemed the highest priority at the time, or by divide attention among equally important sources.

The tasks themselves vary considerably in these jobs in terms of priority, urgency, length of time to complete, cognitive resources demanded, etc. Most of their tasks are performed take seconds to minutes to perform, as predicted. However, some responses take only milliseconds, e.g., some of the LCAC operator tasks that involve control of the LCAC. On the other end of the time scale, division commanders may take several hours to work out a particularly cognitively demanding problem. Again, the time scale the division commander works under allows him to take the time for these kinds of tasks. That said, the division command is still time pressured, as are all of the jobs we studied. In each job, the worker is faced with urgent tasks or he/she has too many tasks to perform in the time available. All of the jobs required extensive knowledge base developed through training, education, and experience.

Feedback about each decision and action made by a worker is notably absent in all of the environments studied, which means that the workers determine how and when they perform each task. It also means that they themselves must decide what constitutes adequate performance.

COGNITIVE OPERATIONS

As shown in Table 4, the four jobs studied in this research varied slightly more in the kinds of cognitive operations they required. The memory requirements they place on workers were very similar. All of the jobs require STM storage of information (e.g., headings for LCAC navigators and operators, vital signs for nurses). LTM retrieval of domain-specific knowledge learned in training or on-the-job experience was also necessary in each of the jobs we studied. Most, but not all, jobs required prospective memory. However, interviewees reported that two jobs do not require prospective memory, probably because they include external sources that provided cues to initiate tasks. For example, the division commander has an aide who helps him keep track of the multiple important demands he faces. The LCAC Operator does not have an aide to help him remember his tasks, but the environment itself is the cue that signals his tasks. The LCAC Operator, in contrast to the Navigator, relies on the environment and his fellow crewmembers to signal necessary actions. Because STM was important to each of the jobs, rehearsal to maintain the contents of STM was as well. However, it is interesting to note that again because of the division commander's aide, STM rehearsal was not a cognitive operation in which he engaged. Updating of working memory was extremely important to all of the jobs. The need to maintain situation awareness, whether one is a combat leader, nurse, chef, or LCAC crewmember, is critical in these dynamic environments.

Table 4.
Summary of Cognitive Operations Required by MT Environments
(√..... cognitive operation used in MT environment)

#	Cognitive Operations	LCAC Operator	LCAC Navigator	Platoon Leader	Company Leader	Division Leader	IC Nursing	Floor Nursing	Chef
---	----------------------	---------------	----------------	----------------	----------------	-----------------	------------	---------------	------

1	Retrospective memory (STM)	√	√	√	√	√	√	√	√
2	Retrospective memory (LTM)	√	√	√	√	√	√	√	√
3	Prospective Memory	Initiate tasks as needed by env.	√	√	√	Has aide for this function	√	√	√
4	Monitoring Output of automatic Responses	√		√	√		√	√	√
5	WM Updating and Monitoring	√	√	√	√	√	√	√	√
6	Mental Set Switching	√	√	√	√	√	√	√	√
7	Classification		√	√	√	√	√	√	√
8	Rehearsal for STM Storage	√	√	√	√	Has aide for this function	√	√	√
9	Selective Attention	√	√	√	√	√	√	√	
10	Divided Attention	√	√	√	√	√	√	√	√
11	Prioritizing		√	√	√	√	√	√	√
12	Deductive Logic		√	√	√	√	√	√	

The control of attention was also critical to performance. In each environment, multiple sources of information were available and were often presented simultaneously. For this reason, workers must decide whether to selectively focus on one piece of information, or divide their attention among several. The relative importance of information seems to be the key determinant whether one takes the strategy of dividing or focusing attention. If the consequences of missing information are severe, one must use a divided attention strategy. If the environment maintains or repeats presentation of the information, or if a particular task cannot be interrupted because it is too cognitively demanding, a selective strategy may be used. All jobs required both selective and divided attention.

The fact that multiple, very different tasks are required by these environments means that (1) workers must switch mental sets when going between tasks and (2) that prioritizing is key to good performance. Indeed, each of our respondents, with the exception of the LCAC operator, reported that prioritization was key. They also reported that it was the hardest element of the job, and took them the longest to learn. As a novice, all tasks seem critical. With time, respondents told us they learned that some are actually more urgent than others. Based on their responses, if there is one factor that determines whether one does well in these jobs or not, it is the ability to prioritize effectively. The LCAC operator didn't report the need to prioritize probably because his tasks were largely cued by the environment.

An interesting outcome of the interviews concerns the need for classification. We originally included classification as a cognitive operation of interest because of the hypothesized lack of feedback in these environments. We reasoned that if the environment did not give adequate feedback concerning the adequacy of a worker's

performance, the worker would have to determine that himself/herself. Hence, we reasoned that the worker would classify either particular actions or the entire performance as either adequate or not. What was interesting was that when we asked the question about whether classification was something they did in their jobs, most respondents replied that it was. However, it had little to do with whether their performance was adequate. The LCAC Navigator, for example, classifies objects as either potential obstacles the LCAC might collide with, or objects that are simply items to note. The nurse classifies the conditions of patients as either moving toward their goals, or not. Combat leaders use classification in identifying enemy units and battlefield patterns. Hence, it appears that classification is integral to many jobs.

Deductive logic was also used in most environments, but not all. In particular, the LCAC Operators and Chefs did not report the need to use deductive logic, except in only the simplest of ways. For example, the chef might use it to determine when dishes should be placed on the burner. The LCAC Operator might think, for example, if there is an obstacle, maneuvering is required. These simple kinds of deductive logic are met with a positive match with one condition. However, these examples do not match the complexity of deductive logic used by combat commanders, nurses or LCAC navigators who must satisfy multiple conditions using deductive logic. Hence, it is not clear to what degree deductive logic is uniformly called upon in MT environments.

It is also important to note that not all jobs apparently require extensive monitoring of automatic responses. This operation probably is more important when the job incorporates a significant proportion of proceduralized tasks. For example, monitoring of automatic responses was important to nurses who must learn to proceduralize a sequence of steps to deliver fluid medication, and who extensively repeat very similar, yet different, tasks (e.g., delivery of pill form medication). Perhaps, when the relative proportion of cognitively demanding to proceduralized tasks is high, as is the case in the LCAC Navigator position, monitoring of automatic output is not as important.

A special note should be made about planning. Although we did not include it in the list of cognitive operations for reasons explained earlier, it did appear to be a cognitive activity important to some, but not all jobs. For example, LCAC Navigators and combat leaders engage in extensive planning before they engage in the MT environment. Similarly, nurses create a plan for each patient at the start of their shift. However, some jobs are inherently more reactive than proactive, like restaurant food preparation. In this case, creation of a detailed plan of action is not possible.

CONCLUSIONS

If we were to design a test of MT ability that would incorporate the cognitive operations most real-world MT environments require, what would it include? Based on the results of our analysis, we propose a test should require that test takers engage in the following cognitive operations.

- STM memory storage
- LTM retrieval
- Prospective memory
- WM updating and monitoring

- Mental Set Switching
- Classification
- STM rehearsal
- Control of attention required by simultaneous presentation of stimuli
- Prioritization

Because it is not clear whether deductive logic is uniformly important, its inclusion in a test remains open to question.

CHAPTER FOUR: ANALYSIS OF EXISTING MEASURES OF MULTI-TASKING

To better understand the cognitive processes and operations that current measures of MT assess, we first conducted a thorough review of the literature to identify measures that other researchers have used. Relevant literatures residing on a variety of databases were searched. The resulting hits were examined for relevance and high payoff sources were obtained. Selected sources were reviewed and pertinent information was extracted about measures of MT.

LITERATURE REVIEW METHODS

A systematic search of the most recent (within the past 5 years) relevant literature was conducted in which a variety of academic and government databases was queried. The following databases were searched for published information concerning multi-tasking.

- ERIC: Educational literature
- NTIC and DTIC: The Military and Federal Government literatures
- PsycINFO: Psychological literature

The Keywords and Title fields of the databases were queried using the relevant search terms such as multitask, multi-task, multitasking, multi-tasking, timesharing, time-sharing, time-pressured decision-making, time pressured decision making, task switching, and executive control AND central executive AND working memory. The names of certain key researchers in the field of multi tasking and related fields were also used to identify their most recent work. These authors include Ackerman, Anderson, Burgess, Kieras, Kyllonen, Meyer, and Pashler. Additional sources were identified from the references sections of reviewed sources; the Internet was also searched for relevant information, focusing primarily on the leading researchers' web sites.

A total of 343 documents were returned from the searches. As is true of any search, the results included hits that were only tangentially related to the topic of interest. Other reports were inappropriate to review for other reasons. Many of the relevant articles had either been discussed in our original literature review or were already in hand. Sixty-five sources were reviewed. A reference list of sources reviewed can be seen in Appendix A of this report.

RESULTS

Researchers have studied MT using various types of measures. One type has been employed to assess neuropsychological disorders; measures involved the application of strategy, planning, and executive control of working memory. A second type has been employed in the simulation of work environments. A third type, stemming from basic research efforts, has addressed the limitations of human performance. Here, the dual- or tri- task paradigm has been used to assess how individuals distribute cognitive, perceptual, and motor resources in laboratory situations that contain multiple simultaneous demands. We begin with a discussion of measures used to assess neuropsychological disorders.

MEASURES DESIGNED TO ASSESS NEUROPSYCHOLOGICAL DISORDERS

The Multiple Errands Test (MET), Six Elements Test (SET), and Greenwich Test have all been employed to assess neuropsychological disorders. Each test is described in this section along with the cognitive operations of each.

Multiple Errands Test. The multiple errands test (MET) (Shallice & Burgess, 1991) was designed to assess executive control dysfunction in brain-damaged patients. While patients who exhibit deficits in this area may perform well on standard tests of neuropsychological functioning, they have trouble in their everyday world. Shallice and Burgess explain the problems as deficits in cognitive control, using Baddeley's model of working memory, particularly the executive control component, as a model. The MET is based on the real world task of shopping. Participants are asked to buy a list of items and are given a limited amount of money to do so. Rules given in the instructions of the test constrain and guide their shopping activity. They must find out certain information, be at a certain location at a specific time and refrain from violating certain rules such as "you must not enter a shop other than to buy something. The MET requires organization, prioritization, and the execution of several different tasks with a given time period. It appears to tap the ability to create delayed intentions and follow them. The MET is scored by noting the tendency to break rules, leave items unfinished, adequacy of plan, failure to carry out planned tasks, and violations of social convention.

The MET has been shown to be a useful measure of dysexecutive syndrome, or what others have called strategy application disorder or frontal lob syndrome, (Baddeley, 1996; Burgess, Veitch, de Lacy Costello, & Shallice, 2000; Wilson, Evans, Emslie, Alderman, & Burgess, 1998). However, it would be of limited use as measure of MT ability or as a predictor of performance in MT environments. First, it is simply not practical to administer because it requires test takers to actually go on a closely-monitored shopping trip. It could not practically be administered to large numbers of people.

Second, it was designed for patients with neuropsychological disorders, not normal populations. Most of the research using the MET has used such patients as subjects. Hence, little is known about how individuals without neuropsychological disorders would score on the test. We suspect that the MET would simply not be demanding of normal populations. Although it appears to tap certain cognitive components of MT demanded by MT work environments, it may not discriminate among individuals from

normal populations at the high end of the distribution of MT ability. We speculate, although this has not been demonstrated, that the MET would produce a ceiling effect for normal populations. The MET, as far as we know, has not been correlated with measures of actual or simulated job performance in real-world MT environments.

Six Elements Test (SET). The SET is a modified version of the MET designed for the laboratory. It constitutes a cognitive analogue of the MET. Subjects are asked to perform up to three open-ended tasks within a 15-minute period of time. Each task is divided into two sections; hence, the three tasks taken twice each form 6 elements. The three tasks that subjects must perform are to describe memorable events, write answers to simple arithmetic sums, and write names of items in simple line drawings. Participants are told that they have 15 minutes to score as many points as they can, but their actions in performing the tasks are constrained by a set of rules. For example, within each section of each task earlier items are given more points than later ones and they are not permitted to perform the first section following by the second section of that same task (Burgess, 1998; Shallice & Burgess, 1991).

Although the laboratory characteristic of the SET makes it a far more practical test of MT than the MET, its predictive utility for normal populations is similarly questionable. The SET has been more often used in research than the MET and control groups of normal populations have taken the SET. Hence, more is known about how normal populations score on the SET. For example, 216 non-brain-injured control subjects, 78 subjects with neurological disorders, and 31 schizophrenic subjects performed the SET in one study of dysexecutive syndrome (Wilson, Evans, Emslie, Alderman, & Burgess, 1998). In this study, normal subjects produced an average score of 3.51 ($SD = .80$) on a modified version of the SET compared to a mean score of 1.99 ($SD = 1.18$) produced by brain-injured subjects. Shallice and Burgess (1991) also showed that three individuals who had suffered brain damage took longer to perform the SET than normals and scored worse. In this case, 10 normal individuals scored an average of 5.7 on the SET. However, these data reveal little about whether the SET could be used to discriminate MT ability in normal populations. As with the MET, the SET has not been used to predict real-world MT performance.

Greenwich Test. The Greenwich Test (Burgess, Veitch, de Lacy Costello, & Shallice, 2000) is an analogue of the SET that requires fewer task switches, but a greater number of rules to follow. It consists of three open-ended tasks that subjects must attempt within a ten-minute period of time. In the first task, the subject is asked to separate green and red plastic beads into two boxes by color. The rules of the task include that the lid of the container that holds the beads must be replaced each time a bead is taken out, and that the beads must be taken out one at a time in an alternating sequence of color. In the second task, subjects are asked to write down letters that label a set of two interlacing lines drawn on paper. The stimuli for the second task consist of two sheets of paper with 10 interlacing lines drawn on each of them. The beginning of each string is marked with a color. At the end of the string, each line is marked with a letter. Subjects must identify the letter of each line at the end of the line that is marked by a color. The third task requires subjects to replicate an object made out of colored pieces of plastic. Subjects must perform these three tasks following a set of rules and scoring constraints.

For example, completing a red item earns a greater number of points than completing an item of any other color.

Like the MET and the SET, the Greenwich test was designed to assess MT deficits in neuropsychological patients. It has been used in conjunction with other behavioral measures to develop a theory of the mental procedures that underlie multi-tasking. A three-construct structural equation model provided the best fit to the data, identifying retrospective memory, planning and intentionality as key latent constructs predicting performance (Burgess, Veitch, de Lacy Costello, & Shallice, 2000). Retrospective memory is the ability to remember information and tasks already completed. Planning is the ability to develop an organized plan that dictates sequential and related tasks. Plans must then be followed for performance to be successful. Intentionality is the ability to create and remember future tasks, and is what is often called prospective memory. The three-factor structural equation model was developed using various subscores or measures of the Greenwich obtained from 90 patients and 60 controls. However, the applicability of the model and the Greenwich test to normal populations is in question. At this point, little is known about (1) the ability of any of these tests to discriminate among high levels of MT ability, and (2) the ability of the tests to predict actual or simulated performance in real-world MT environments.

Cognitive Operations of MET, SET, and Greenwich. Burgess et al. (2000) argue that prospective memory is the cognitive component that distinguishes a test like the SET, and we assume analogous tasks like the MET and the Greenwich test, from other experimental tasks that involve the concurrent processing of multiple tasks. Other tasks certainly require retrospective memory and planning, but they don't necessarily require prospective memory, which is the realization of delayed intentions. They argue that individuals working in naturalistic MT environments must decide for themselves what goals to set and to determine when they have reached those goals. Hence, the environment does not provide external signals that guide or even force the individual to perform tasks in any particular sequence or by any particular strategy. Other types of laboratory tasks (e.g., most versions of the PRP, task switching, and many dual task situations) give explicit instructions regarding task priorities and scheduling or afford a particular sequence of task execution. Burgess et al. (2000) argue that MT is more than task switching or simple task interference as found in the PRP procedure because these tasks do not involve the deferral of task execution over lengthy periods of time. They also note that MT environments in the real world typically involve how attentional resources are allocated to competing demands.

Burgess et al. (2000) propose that the SET (and by analogy the MET and Greenwich) requires the test taker to employ the following cognitive operations.

1. *Retrospective memory (STM)*

- Test takers must remember the rules of the task
- Test takers must be able to learn the rules of the task
- Test takers must remember the tasks they've already accomplished
- Test takers must remember the plan and the list of items to buy

2. Prospective memory

--Test takers must be able to remember the plan and its components

They also propose that planning is a third cognitive operation incorporated by these tests. However, analysis of SET data taken from 150 normal and brain-damaged patients to test a three-construct structural equation model in Burgess et al. research failed to distinguish it from a two-construct model. They were not able to demonstrate that planning is a separate cognitive operation from prospective memory, although anatomical data taken in the same study did support a three-construct model that included planning. We would add to Burgess' list of three cognitive operations other cognitive operations (given below) required by the MET, SET, and Greenwich tests. It is also important to note that these tests do not require divided or selective attention to be allocated among simultaneously presented stimuli. The kinds of interruptions that normally demand control of attention are not present in the MET, SET, and the Greenwich tests. The MET shopping task draws upon CTM stored knowledge, but the SET and Greenwich tests do not. Test takers do not have to prioritize tasks nor do they use deductive logic for any of the subcomponents of the tests.

3. *Monitoring output of automatic processes, inhibition of prepotent responses*

--Test takers must inhibit the actions of inappropriate behaviors and attention to nonrelevant streams of information. Although Burgess et al. do not consider inhibition as a cognitive operation in these tests, other research has shown that brain damaged patients exhibited inappropriate behaviors when taking the MET (e.g., climbing on displays in store windows).

4. *WM Updating and Monitoring*

--Test takers must monitor their behavior and update their WM in terms of tasks accomplished, progress achieved, in order to know whether they are meeting their goals, intentions, and plans.

5. *Mental Set Switching*

--Because the tasks are quite different in all three tests, participants must change mental sets when switching among tasks.

6. *Classification*

--The Greenwich test requires test takers to sort beads based on color

7. *Rehearsal for STM*

--Memorization of rules may require rehearsal

MEASURES DESIGNED AS SIMULATIONS OF REAL-WORLD WORK ENVIRONMENTS

Three measures have been designed to simulate real-world work environments—SYNWORK, Multiple-Attributes Test Battery (MATB), and the Abstract Decision

Making (ADM) task. The three are described and discussed and their cognitive operations identified in this section.

SYNWORK. SYNWORK is a "synthetic work environment" (Proctor, Wang, and Pick, 1998) designed to afford laboratory investigation of human performance in real-world work environments. It is a PC-based system that includes four component tasks concurrently displayed in four quadrants of a computer monitor. The lower half of the monitor displays two tasks that require visual or auditory monitoring. In the visual monitoring task, the subject's task is to recognize when a marker displayed on a horizontally oriented scale moves either to the right or left of center. Clicking on the RESET button via a mouse click returns the marker to the center of the scale. In the auditory monitoring task, a series of tones are presented periodically. The subject's task is to click on a HIGH SOUND REPORT button in the lower right hand quadrant of the screen when a tone is higher in pitch than normal. The upper half of the screen displays a memory retrieval task and an arithmetic task. In the memory task, the subject first memorizes a list of letters that constitutes the memory set. A single letter is then periodically presented and the subject must decide whether the letter is included in the set by clicking a YES or NO box in that quadrant of the monitor. For the arithmetic task, two two- or three-digit numbers are displayed and the subject's task is to adjust a third set of numbers to equal the sum of the first two. When a DONE button is clicked, new numbers appear. Performance is scored in the memory retrieval and arithmetic tasks on the basis of accuracy. Points are earned or lost on the basis of whether the subject makes a correct or incorrect decision, respectively. Points are earned on the monitoring tasks based on how far the marker gets away from the center of the scale in the visual task, and the length of time it takes to respond to a pitch that is out of "normal". Total points earned for all tasks are displayed in the center of the screen and is continually updated. The computer-based format of SYNWORK affords flexibility and utility to investigations of factors that affect work performance. The tasks can be performed singly or in any combination. The difficulty of the component tasks can also be varied, as can the payoff matrix.

While SYNWORK is a valuable experimental laboratory tool for investigating factors that influence performance in work environments under single and multiple task conditions, its ability to predict multi-tasking ability and MT performance in simulated work environments or actual work environments has not yet been assessed. It is also limited because the work focuses on visual monitoring, which requires vigilance because heavy penalties are applied when items go unnoticed, and arithmetic, which is the most cognitively demanding of the four tasks. The scoring emphasis on visual monitoring and arithmetic may or may not have ecological validity to real-world MT environments. At this point in time, little is known about the specific tasks common to MT environments. Modifications to SYNWORK could certainly be made to include tasks that demand cognitive operations required by most MT environments. However, to our knowledge research using SYNWORK to predict MT ability has not been conducted. We discuss the similarities between SYNWORK and real-world MT environments later in this section of the report.

Multiple-Attribute Task Battery (MATB). The MATB is a computer-based synthetic task battery that can be used in the laboratory to simulate MT specific to aviation.

Similar to SYNWORK, it displays multiple tasks on a computer screen. The MATB, however, uses tasks that are analogous to aviation tasks such as tracking, system monitoring, fuel management, and communications. The MATB is a flexible laboratory tool originally designed to investigate issues of aviation workload (Comstock & Arnegard, 1992). For example, in one use of the MATB the system monitoring task required attention to 4 gauges and two boxes where the subject could manipulate the boxes and gauges by pressing keys. Pointers in the gauges varied between the desirable one tick mark above or below a mid-line. When the subject pressed a key, pointers that are beyond the mid-line returned to the correct range. The tracking task required keeping a target in the center of a window using a joystick. The communications task simulated reception of radio messages from Air Traffic Control. Here, the subject was asked to make appropriate frequency changes on the radio and discriminate their own call sign (three letter or three number combinations). The fuel management task was similar to the system monitoring task in that the subjects were required to maintain the tanks at a certain level by turning a set of pumps on or off with key strokes. A pump failure sometimes occurred, in which the pump turned red indicating it could not be used. This task also allowed the subject to transfer fuel by activating pumps. Each task of the MATB can be fully or partially automated to investigate factors that influence performance in aviation.

The MATB has been shown to require a high level of cognitive resource sharing and has been rated by subjects as a good face-valid method for assessing aviator performance (Caldwell & Ramspott, 1998). It has been used to investigate a variety of factors related to aviation performance such as sleep deprivation (Caldwell & Ramspott, 1998), self-regulation to monitor task engagement (Prinzel, Pope, & Freeman, 2001), automation-induced complacency (Prinzel, DeVries, Freeman, & Mikulka, 2001), and the effects of unreliable automation on aviation workload (Rovira & Zinni, 2002), among other topics. The MATB has also been modified to cover Army Infantry scenarios, producing a derivative synthetic MT environment called "Viking" (Harris, Parasuraman, Zinni, Hancock, & Harris, 2002).

Like SYNWORK, the MATB is a very useful tool for investigating issues related to work performance in MT environments. However, because of its focus on aviation tasks, its ability to predict performance in jobs other than aviation (or Infantry for Viking) is questionable. Like SYNWORK, there is no reason to believe that the particular combination of tasks used in MATB generalize to *most* MT environments.

Cognitive Operations of SYNWORK and MATB. In this section, we document the cognitive operations SYNWORK and MATB appear to incorporate, or not, as well as the evidence we use to make that judgment.

1. *Retrospective Memory (STM)*

- Test takers must remember the list of letters in the memory set in SYNWORK.
- Test takers must remember the target values of the monitoring tasks in both the SYNWORK and the MATB

- Test takers must remember their call sign on the MATB
- 2. *Retrospective Memory (LTM)*. Long term memory retrieval of well learned domain specific knowledge is not required in SYNWORK. However, MATB can only be taken by individuals who have some knowledge of aviation.
 - Test takers must monitor and manage fuel level on the MATB
 - Test takers must receive simulated ATC messages on the MATB
- 3. *Prospective Memory*
 - Test takers must remember to perform the tasks contained in SYNWORK and MATB
- 4. *Monitoring Output of Automatic Processes*
 - Test takers must monitor their response to the memory search task of SYNWORK to determine if the presented letter is in the memory set
- 5. *WM Updating and Monitoring*
 - Test takers must update their WM of the status of visual and auditory monitoring tasks
 - Test takers must update their WM of the status of the system, resource management, and communications tasks
 - Test takers must update their WM of the memory set when a new one is presented in SYNWORK
 - Test takers must update their WM of a new letter presented in the memory search task in SYNWORK
 - Test takers must update their WM of the arithmetic task when a new equation is presented in SYNWORK
- 6. *Mental Set Switching*
 - Because the tasks are quite different in all tasks of SYNWORK and MATB, participants must change mental sets when switching among tasks.
- 7. *Classification*. *Classification is not required by the MATB*
 - The memory set task in SYNWORK requires test takers to classify each target letter
- 8. *STM Rehearsal*
 - Test takers may rehearse memory set in SYNWORK.
 - Test takers may rehearse target values of the monitoring tasks in both the SYNWORK and the MATB
 - Test takers may rehearse their call sign on the MATB

9. *Selective Attention*

- Test takers must choose to attend to one set of incoming information from the visual and auditory monitoring tasks in SYNWORK, especially when they are out of range because severe point losses are incurred when their response to these tasks are delayed
- Test takers must choose to attend to each of the four tasks of the MATB

10. *Divided Attention*

- Test takers may divide their attention between the four tasks of SYNWORK, especially between the two monitoring tasks
- Test takers may divide their attention between the four tasks of the MATB as all require some level of monitoring

11. *Prioritizing*

- Test takers must prioritize among their possible responses because points earned on the monitoring tasks in SYNWORK are severely reduced if response is delayed. Conversely, points earned on the arithmetic task are high. Hence, some prioritization strategies earn more points than others.
- Test takers must prioritize among the four tasks in the MATB depending on the points earned for each task, which can vary as determined by the particular experiment

12. *Deductive Logic. There is no substantial deductive logic requirement on SYNWORK or the MATB*

Abstract Decision Making (ADM). Joslyn's and Hunt's (1998) ADM task was developed as an abstract version of public safety dispatching, which involves the allocation of limited resources in the performance of a fundamental classification operation. Emergency dispatchers must assign resources, such as police or fire units, to situations based on classification of each case. They may be required to simultaneously handle several cases at any one point in time, which involves not only making the appropriate resource allocation decision, but also monitoring the progress of each situation. Likewise, air traffic control (ATC) also involves the allocation of limited resources, in this case airspace, to multiple cases based on classification. Hence, in dispatching and ADM, classification is fundamental to the job and it is made using partial information about the attributes of a stimulus.

The ADM task (Joslyn & Hunt, 1998) was designed as an abstract task that had the cognitive elements of decision-making in an MT environment, but lacked the specific content of any particular MT environment. Joslyn's and Hunt's purpose in developing this task was to predict performance in a variety of time-pressured decision-making situations. They sought to determine if individual differences in MT ability are due to the specific demands of particular jobs, or a general ability to make decisions in time-pressured situations.

ADM is a computerized task that is largely text based, but has the feel of a computer game. Subjects earn points by making sorting decisions about objects. The objects are not actually shown to the subject. Rather the subject is informed by a text message that an object is available for examination. The subject may then ask a series of questions about the object concerning its shape, size, or color. Classification of the objects is based on a set of bins, or categories, that the subject is presented with at the start of the game. The bins describe characteristics of objects that "fit" in the bins. For example a bin could be described as a red object of any size or shape. Or a bin could be described as a small blue triangle. The number of bins can vary, but in Hunt's and Joslyn's experiments three to four bins were presented to subjects. As Hunt and Joslyn note (2000), this description of ADM makes it sound exceedingly easy. However, nothing could be farther from the truth. The task qualifies as an MT environment because objects have a 50% probability of being made available to the subject every 15 seconds in the experiments reported, although this speed can be varied. Availability of a new object typically occurs before classification of a previous object. Hence, subjects are "working on" multiple tasks at any particular point in time because they have multiple objects to classify. Objects in ADM are identified by number (e.g., #9), and the system requires that the subject specify which object he/she wants to query or classify. So the subject must remember the object numbers that are currently available, whether or not they have been classified, what characteristics they hold, and what characteristics have been queried.

From a practical perspective, ADM has many attributes that make it a good candidate for a test of MT ability and prediction of performance in MT environments. First, it can be easily administered. It currently takes about 30 to 45 minutes for a subject to take the ADM in the form it was used by Hunt and Joslyn in their experiments. However, it is likely that the test could be reduced and still enjoy the same high levels of psychometric reliability and, hence, potential for predictive validity.

ADM also has the benefit of having been designed to predict MT in real-world environments without the trappings of specific topics or tasks idiosyncratic to particular jobs such as ATC, aviation, etc. It is abstract in nature, as Joslyn and Hunt attempted to make it a general measure of MT ability.

Perhaps the characteristic that makes ADM the current best candidate measure of MT is that it has been demonstrated to predict simulated job performance in three very different MT environments (emergency call answering, dispatching, and ATC) at unusually high levels of predictive power.

Despite its apparent high utility as a test of MT, several questions still surround its applicability. While the ADM task is highly correlated with performance on simulated versions of dispatching and ATC jobs, it may not predict performance of other real-world multi-tasking jobs. Joslyn and Hunt (1998) acknowledge that it is possible that the ADM task has limited universal generality and note that research is needed comparing the ADM task to other multi-tasking jobs such as medical emergency and tactical decision-making in military situations. Second, while Joslyn and Hunt were able to predict performance on laboratory simulations of 911 dispatching and ATC, they have not demonstrated that the ADM task predicts actual performance on the job.

Moreover, the cognitive components that ADM requires may or may not overlap with MT environments other than ATC and dispatching. The following section describes our analysis of ADM's cognitive components.

Cognitive Operations of ADM

1. *Retrospective Memory (STM)*
 - Test takers must remember which items they've classified
 - Test takers must remember the attributes of each bin
 - Test takers must also remember the known attributes for each item
 - Test takers must remember which item they were querying
2. *Retrospective memory (LTM)*. Long term memory retrieval of well learned domain specific knowledge is not required in ADM
3. *Prospective memory*
 - Test takers must remember which items must still be classified
4. *Monitoring output of automatic processes.*
 - Test takers must inhibit memory and response to bin content of previous session
5. *Working Memory Updating*
 - Test takers must update attribute information on each object after each query
 - Test takers must update fit to the bins on each object after each query
 - Test takers may update their memory of the current bins' attributes by querying the bins
 - Test takers update their memory of object status each time they assign an object to a bin. This allows the test taker to "drop" memory of that particular object.
6. *Mental Set Switching*
 - Test takers must alternate between querying object attributes, classifying objects, querying bin contents, specifying object #, etc.
7. *Classification*
 - Test takers must assign objects to a bin based on the match between object attributes and bin attributes. In this task, classification is the process of making a decision. Hence, we do not include a separate cognitive operation of decision-making, which may well be at a different level of abstraction anyway.

8. *Rehearsal for memory storage*

--Test takers must rehearse the numbers of those items not yet classified as well as item attributes to ensure that they are kept in short-term memory.

9. *Selective Attention*

--Test takers must inhibit attention to some tasks so that others may receive focus. For example, it may be necessary or desirable to pay minimal attention to new items arriving on the screen when querying or classifying another item. Conversely, it may be desirable to attend to the number of the new item that has been made available, while inhibiting attention to the item one had been processing when the new item arrived.

10. *Divided Attention*

--Test takers may divide their attention between newly available objects and the object they are querying or classifying, or the bins they are querying at any point in time.

11. *Prioritizing*

--Test takers must decide which subtask (query, assign, look at bins) to perform first. Subjects must also decide which item to classify.

12. *Deductive Logic*

--Test takers must set up a strategy for querying objects using deductive logic based on the attribute contents of the bins. For example, bin attributes may permit querying of a single attribute to determine assignment. Individuals who use deductive logic pertaining to the relative attributes of the bins may use this knowledge to efficiently query the system and then assign objects.

MEASURES DESIGNED TO INVESTIGATE THE LIMITS OF HUMAN INFORMATION PROCESSING

Extensive research on human information processing has been conducted using the dual-task paradigm. The various measures involved in this research are described in this section along with the cognitive operations of dual tasks. In addition, measures involving information coordination tasks and the psychological refractory period (PRP) measure are described and discussed together with the cognitive operations involved in each.

Dual Task Paradigm. The dual-task paradigm has been widely used in the laboratory to investigate limitations on human information processing. Earlier studies used the dual task paradigm to test theories about the allocation of attention and cognitive resource models (e.g., Wickens, 1980; Kahneman, 1973). Recent studies have used the paradigm to test models of working memory paying particular attention to the central executive component of Baddeley's theory of working memory (e.g., Emerson, & Miyake, 2003; Hegarty, Shaw, & Miyake, 2000; Miyake, Friedman, Emerson, Witzki, Howerter, and Wager, 2000).

By definition, the dual tasking laboratory paradigm is a kind of multi-tasking. Therefore, we consider it here as a potential measure of MT ability. Note, however, that the literature is replete with variations of dual, and sometimes, triple, tasking laboratory tasks. Researchers have combined any number of combinations of verbal, visual, spatial, or auditory tasks that may vary in other ways as well. For example, tasks may be selected because they are thought to be relevant to particular functions of the central executive (Emerson & Miyake, 2003). Or tasks may be selected because they are thought to demand different cognitive resources (non competitive) or the same cognitive resources (competitive). Relatively little is known about how well dual tasks predict MT ability in real world MT environments as most of the research has focused on basic research questions using experimental, as opposed to predictive or correlational, methods (a notable exception is Gopher & Kahneman, 1971). That said, the clear potential for the use of dual task situations for predicting MT ability leads us to discuss and consider those dual task measures we regard as the most relevant to the issues surrounding MT, such as working memory, task switching, and executive or central control of performance.

Personality and Multi-tasking. With the purpose of investigating the relationship between Type A behavior pattern (TABP) and performance in multi-tasking situations, Ishizaka, Marshall, and Conte (2001) developed a computerized test in which three tasks were presented simultaneously. Two of the tasks presented visual stimuli while the third presented auditory stimuli. The first visual task was a math task in which subjects were required to evaluate two mathematical expressions and decide whether the expressions held the same or a different value. Participants responded by clicking on "Same" or "Different" buttons. The second visual task presented six gauges on the left side of the monitor. Each gauge displayed an arrow and an area consisting of red and white zones. The task was to keep the arrow in the white zones. Participants clicked on a button to the right side of each gauge, which changed the direction of the arrow. The auditory task required participants to pay attention to words they heard during the session. Fifteen words were presented, one every 30 seconds. Subjects were required to recall the words immediately after the session.

Separability of Executive Functions. Miyake, Friedman, Emerson, Witzki, Howerter and Wager (2000) investigated whether three central executive functions often discussed in the WM literature are truly separable abilities. In an individual-differences study, they used a latent variable analysis to examine the relationships among mental set shifting, information updating and monitoring, and the inhibition of prepotent responses. Subjects completed fourteen separate tasks, one of which was a dual-task situation. In that situation, subjects were required to perform a spatial scanning task and a word generation task under single and dual task conditions. In the spatial scanning task, a maze tracing speed test, they were required to trace as many mazes as possible within a three-minute period with instructions to avoid retracing any lines or removing the pencil from the paper. In the second task, a word generation task, subjects listened to the presentation of a letter every 20 seconds and were required to generate as many words as possible that began with that letter until the next letter was presented.

The Role of Inner Speech in Task Switching. To identify the role of inner speech in task switching, Emerson & Miyake (2003) required subjects to perform simple arithmetic operations on lists of two-digit numbers. Participants performed the same operation (e.g., addition) to all numbers on some lists and alternated between different operations on other lists. As a second task, subjects also performed either an articulatory suppression task, which involved repeatedly saying 'a, b, c' aloud, or a foot tapping task, which involved repeatedly tapping one's foot. Various combinations and modifications of these tasks were used in four experiments.

Dual-task Methodology and the Central Executive. To investigate the limits of the applicability of dual-task methodology to study of the central executive, Hegarty Shaw & Miyake (2000) presented three visuospatial psychometric tests (taken from the Ekstrom Kit of Factor-Referenced Cognitive Tests) to subjects. These included the paper folding test, the card rotations test, and the identical pictures test. Subjects were also required to perform several secondary tasks, depending upon the condition of the experiment. A random number generation task required participants to generate random numbers at a rate of 1 per second to a the beat of a metronome. In a second secondary task, subjects were asked to listen to a series of consonants displayed at one every two seconds. They were instructed to say "yes" if a consonant was identical to the consonant presented two items before and "no" to all other consonants. The third secondary task was a spatial tapping task in which subjects tapped a square spatial pattern around a numerical keypad by tapping the numbers 1,4,7,8,9,6,3,2 in that order.

Cognitive Operations of Dual Tasks. Depending on the particular tasks chosen, the cognitive requirements of the dual-task paradigm can vary. However, despite the fact that the studies we reviewed in this report used a wide variety of tasks, they appear to require a very similar set of cognitive operations. Moreover, the cognitive operations tapped appear to closely match those of SYNWORK and MATB. One exception is that most dual task paradigms require only two tasks, whereas SYNWORK and MATB require at least four. Hence, the requirement of prospective memory is much reduced and even eliminated in many dual task experiments. This is particularly true when the instructions or the task characteristics limit the subject's choices of which tasks to perform. For example, most task switching tasks cue the subject to alternate between two operations, which does not afford subject generated prioritization and does not require the subject to employ prospective memory. Hence, notice that prospective memory has been excluded from the list of cognitive operations required by dual task paradigms.

1. *Retrospective Memory (STM) (Not all dual-tasks require retrospective memory, but many do)*
 - Test takers must remember the words they heard in the auditory task
 - Test takers must remember the letter presented in the word generation task
 - Test takers must remember previously presented consonants
 - Test takers must remember the appropriate sequence of tapping on a numerical keypad

2. *Retrospective Memory (LTM).* Long term memory retrieval of well learned domain specific knowledge is not required in most dual-tasks
3. *Prospective Memory.* Prospective memory is not required in most dual-tasks because tasks are cued by instructions or environment.
4. *Monitoring Output of Automatic Processes, inhibition of prepotent responses*
--Test takers must monitor the suitability of responses that they generate
5. *WM Updating and Monitoring*
--Test takers must update their WM of the status of visual and auditory monitoring tasks
--Test takers must update their WM when a task presents a new stimuli to work on, such as a new math equation or a new letter for a word generation task
6. *Mental Set Switching*
--Because the tasks are usually quite different, participants must change mental sets when switching among tasks.
7. *Classification.* The need for classification depends on the particular tasks selected for the dual-task paradigm
8. *Rehearsal for Memory Storage*
--Depending on the particular set of tasks, test takers may or may not have to rehearse stimuli presented in an auditory or visual mode
9. *Selective Attention*
--Test takers must choose to attend to one of the tasks at a time, particularly monitoring tasks
10. *Divided Attention*
--Test takers may divide their attention among or between the tasks
11. *Prioritizing*
--Test takers must prioritize according to instructions or devise their own prioritization according to how performance is measured or preference
12. *Deductive Logic.* Deductive logic is typically not required by dual-tasks, but may be depending on the particular tasks chosen in the paradigm

Information Coordination. According to Yee, Hunt and Pellegrino, (1991) information coordination tasks differ from dual tasks, or multiple tasks, in that dual task performance can be explained by resource competition models (e.g., Kahneman, 1973, Norman & Bobrow, 1975) whereas coordination performance requires the integration of

the products of the multiple tasks. Hence, resource competition models may not represent the kinds of mental processes used in information coordination tasks. They note that the real world of multi-tasking involves coordination at least as much as it involves resource sharing among multiple different tasks. For example, pilots must coordinate reports they receive from ATC with information from their instruments. Basketball players must coordinate visual information about the placement and movement of own and opposing team members with their coach's instructions and their knowledge of common and practiced patterns of play. Yee Hunt and Pellegrino (1991) argue that the coordination of multiple sources of information is a task in-and-of itself. Hence coordination tasks are inherently more complex than multiple task situations that do not require the integration of information.

However, other researchers have demonstrated that performance in coordination tasks is correlated with performance in multiple, but unrelated, task situations (Emerson, Miyake, and Rettinger, 1999). Emerson et al. also found that both kinds of performance were correlated with the ability to switch attention. Hence, the predominant factor that influences individual differences in IC tasks may not be the ability to integrate information as Yee, Hunt, and Pellegrino first hypothesized. Instead, it may involve the executive function of switching between tasks, which would be common to both related and unrelated sets of tasks.

The laboratory paradigms that have been used to investigate IC have typically employed tasks in which participants must integrate verbal information with related visual-spatial, or with auditory, information. For example, Yee, Hunt, and Pellegrino presented subjects with a task in which subjects were asked to determine which of two objects would arrive at their respective destinations first. In the dual task condition, subjects were simultaneously presented with a verbal statement that made the proposition that one of the objects would arrive before the other. Grammatical complexity of the verbal statement was manipulated to control difficulty. Subjects were required to determine whether the statement was a true or false description of the visually presented task.

Cognitive Operations of IC Tasks. Note that retrospective LTM or STM, prospective memory, deductive logic, rehearsal, and prioritization are typically not required in IC tasks.

1. *Retrospective STM.* STM storage is not required in IC tasks
2. *Retrospective LTM.* LTM retrieval of domain specific knowledge is not required in IC tasks
3. *Prospective Memory.* Prospective memory is not required in IC tasks
4. *Monitoring output of automatic processes, inhibition of prepotent responses*
--Test takers must monitor the suitability of responses that they generate
5. *WM Updating and Monitoring*
--Test takers must update their WM of the status of spatial situation

6. *Mental Set Switching*

--Although there is only one response required and subjects must integrate the information derived from the spatial presentation and the verbal presentation, they must also switch between the two modes to derive a conclusion about each, which must then be compared.

7. *Classification*

--Test takers must make judgment whether spatial presentation matches verbal description

8. *Rehearsal for STM. IC tasks do not require STM rehearsal*

9. *Selective Attention*

--Test takers must choose to attend to the spatial information or the verbal information

10. *Divided Attention*

--Test takers may divide their attention among the spatial information and the verbal information

11. *Prioritization. Prioritization is not required in IC tasks*

12. *Deductive Logic. Deductive logic is not required in IC tasks*

An Elemental Measure of MT: Psychological Refractory Period. An experimental paradigm called the psychological refractory period (PRP) procedure has been used extensively in laboratory studies of the concurrent performance of multiple tasks (Meyer & Kieras, 1997). In this procedure subjects are presented with a series of trials in which two stimuli are presented in sequence. Their task is to make a response to the first stimulus and to make a response to the second stimulus, as well. The time between presentation of the first and second stimuli is the stimulus onset asynchrony (SOA), which typically is varied between 0 and 1 second. This paradigm constitutes, perhaps, the simplest MT environment in which perception of and response to the first stimulus constitutes the first task, and perception of and response to the second stimulus constitutes the second task. The reason the PRP procedure has been widely used by researchers is that it produces a phenomenon, known as the "PRP effect" where response times to the second stimulus are greater than response times to the first stimulus. Moreover, response times to the second stimuli are longer the closer in time the two stimuli are presented. The PRP effect can disappear at longer SOAs. Hence, it appears that the first stimulus-response task interferes in some way with execution of the second stimulus-response task when the SOA is short. The PRP effect has been studied extensively by researchers concerned with the human information processing system's capacity and architecture. Over a 40-year history, the PRP procedure and other attention and performance phenomenon have inspired development of a host of theories that attempts to explain the interference observed in the performance of multiple tasks (Meyer & Kieras, 1997).

The PRP procedure, and corresponding effect, may be considered an elemental, but powerful, measure of MT performance that may well underlie decrements in real-world MT performance. However, it is probably too simple of a task to make it a serious candidate as a predictor of real-world MT ability or performance. Consideration of the cognitive operations most likely tapped by the PRP procedure reveal its probable limitations.

Cognitive Operations of PRP. Note that because of the short duration of the typical response on a PRP trial, the PRP procedure appears to use sensory stores rather than STM. Hence, we exclude STM and rehearsal as a cognitive operation used by PRP. We also exclude selective and divided attention because the cue stimuli are not presented simultaneously. Prospective memory is also excluded because the PRP procedure requires only simple responses to stimuli, i.e., the stimulus itself is the cue to respond. However, subjects can prioritize their responses to the two stimuli. In some cases, the prioritization is instructed by the experimenter. In other cases, the instructions give the subject the choice to set priorities. Classification and deductive logic are also not required in this PRP procedure.

1. *WM Updating and Monitoring*

- Test takers must update their WM that each stimulus has been presented and responded to

2. *Mental Set Switching*

- Test takers must switch between responding to the first and second stimuli, which may require different kinds of responses

3. *Prioritizing*

- Test takers must prioritize according to instructions or devise their own prioritization according to how performance is measured or preference

4. *Monitoring output of automatic processes, inhibition of prepotent responses*

- Test takers must monitor the suitability of responses that they generate

CHAPTER FIVE: GAPS IN THE MEASUREMENT OF MT

COMPARISON OF MT MEASURES

Laboratory tasks (Dual task paradigm, IC tasks, and PRP procedure), which have been extensively and successfully used to examine the fundamental limits of cognition, do not adequately represent the complexity of real-world MT environments in terms of the cognitive operations they demand. First, they typically do not require prospective memory, which is critical to successful performance in the real-world MT jobs we analyzed. Second, while many of the jobs we analyzed required the continuous storage of information in STM, STM rehearsal, and LTM retrieval, these elemental tasks place little demand on these forms of memory and instead rely on iconic or auditory storage. Third, they do not assess more important complex and demanding cognitive processes used in real-world MT environments such as planning and deductive logic. Finally,

while these MT measures do require the participant to prioritize among tasks, we believe that they demand only the simplest kind of prioritization, which does not adequately represent the complexity of real-world MT environments. Prioritization in the MT jobs we analyzed involved knowledge stored in LTM derived from years of experiencing the consequences of inappropriate prioritization schemes. It also involved updating and reorganizing priorities as the situation changes. In summary, IC tasks, the PRP procedure, and most dual tasks, primarily assess the ability to switch tasks efficiently and control attention. However, real-world MT environments are far more demanding and require the use of different kinds of cognitive processes. Perhaps this is why laboratory tasks have been relatively unsuccessful at predicting more complex real-world performance (e.g. IC tasks and WM tasks in Joslyn & Hunt; 1998; Yee, Hunt & Pellegrino, 1991).

Measures developed to assess neurological problems, such as dysexecutive disorder, also fail to adequately represent the cognitive components of MT jobs. While the MET, SET, and Greenwich tests do assess cognitive operations such as setting and following a plan, retrieving information from LTM, storing and using information in STM, remembering future tasks (prospective), and switching among different tasks, they do not present a situation in which a person must divide attention among *simultaneously* presented multiple sources of information nor do they require selective attention. In each of the jobs we analyzed, the need to divide or select attention was a salient and critical component of the environment. Indeed, it is part of what creates an MT environment because the worker cannot control when he or she will receive information. Interruption in dynamic environments, as we have previously discussed, is a defining characteristic of an MT environment because it does not allow the worker to control the sequence of work. As is true of basic laboratory tasks, these neuropsychological measures also do not represent the complexity of prioritization and deductive logic found in real-world MT jobs. Moreover, as we have previously noted, it is highly likely that ceiling effects, or at least range restrictions, would be found in normal populations who take the MET, SET, and Greenwich tests.

Perhaps it is not surprising that the tests that have been purposely designed to simulate or predict performance in real-world jobs appear to best represent the cognitive operations we believe those jobs demand. The SYNWORK, MATB, and ADM tasks all have divided and selective attention components. They all require STM and WM processes such as rehearsal, storage, and updating. They all demand superior ability in prospective memory. MATB can be ruled out simply because it is specific to one field, aviation, making it a test that would most likely not generalize to other domains. Hence, of the existing measures of MT analyzed in this research, SYNWORK and ADM appear to be the best candidates on which to base a test of MT ability. Our analysis indicates that they assess most of the cognitive components required by the eight MT jobs analyzed in this study.

If choosing between SYNWORK and ADM the immediate obvious choice would be ADM if for no other reason than it has already been demonstrated to predict simulated and actual job performance at a surprisingly high level of accuracy. This empirical reality is no small consideration as it is highly unusual to obtain the level of predictive

power that has been demonstrated with ADM. There is no real need to consider the capabilities of SYNWORK given this advantage of ADM.

However, there are other compelling reasons to base a test of MT ability on ADM. First, ADM includes the critical feature of unpredictable interruption, which SYNWORK really doesn't. As one is focusing on querying or classifying an object, ADM presents another available object to classify, which amounts to an interruption that cannot be ignored. Nor can one predict, let alone control, when a new object will be made available. Hence, ADM requires selective, or at least divided attention, so that the number of the object can be encoded, rehearsed, and stored in STM. Knowledge of an object's number designation is critical because further processing of the object depends on that knowledge. ADM requires reference to the number for any action taken on the object. If the number designation is not known, or cannot be guessed, all is lost for that particular object. In the following section of this report we discuss this issue in greater depth because, in fact, the current version of ADM affords the ability to deduce object number designations.

While SYNWORK requires concurrent performance of multiple tasks, it does not incorporate unpredictable interruptions. SYNWORK presents a visual monitoring task that can be scanned at any time. The visual information doesn't disappear, although delay in response may lower points earned. SYNWORK also presents an auditory monitoring task, which must also be responded to within a period of time or points are deducted. However, there is no requirement to respond immediately to the change in pitch that cues the response. Similarly, the arithmetic task and the memory retrieval task can be done whenever the participant chooses. If the test taker forgets the letters stored in STM, he/she can view them again by clicking on a "Retrieve List" button. In short, whereas ADM has an interruption component that cannot be ignored (much like an obstacle for a LCAC Navigator), SYNWORK really doesn't. While modifications could certainly be made to SYNWORK so that it included unpredictable interruption, there appears to be no need to do so since it is already present in ADM.

One might make the argument that SYNWORK includes *different* tasks, and hence represents real-world MT environments better than ADM. However, we believe that this is a specious argument because ADM does require changing from querying objects to classifying objects to querying bins, to encoding a new object number. Alternating among these various tasks require a mental set shift, as does alternating between different, but related, tasks in real-world MT jobs. One advantage SYNWORK may have over ADM is that the tasks it incorporates are a better match to many MT environments. For example, SYNWORK includes visual and auditory monitoring tasks, which are found in many MT environments. However, not all MT environments require visual and auditory monitoring, which may make SYNWORK's selection of tasks inappropriate for some domains. Moreover, in choosing between ADM and SYNWORK, ADM's demonstrated and impressive ability to predict simulated job performance more than outweighs the potential advantage SYNWORK may have because of its selection of tasks.

ADM also includes a significant deductive logic component, which may increase or decrease its predictive utility depending on the particular job in question.. High levels

of performance are achieved in ADM by first figuring the logical structure of the object attributes contained in the bins. The bin structure can be analyzed to deduce the most expedient querying strategy, which can then be used during conduct of the test. Deductive logic is a cognitively demanding task that LCAC Navigators use in re-computing navigational plans when obstacles require course changes. Similarly, deductive logic is used to determine patient status, which is then used to guide future actions of ICU nurses. On the other hand, deductive logic may not be a component of all MT jobs. Individuals who participate in SYNWORK may use deductive logic to determine which tasks are priorities based on their relative point rewards. However, the conduct of SYNWORK tasks does not require deductive logic. The relative utility of a deductive logic component in a test should be evaluated.

In conclusion, current knowledge of MT and its measurement strongly suggest that the best existing candidate for predicting MT ability is ADM. The goal to develop an assessment test of MT ability would be best reached by basing the test on ADM. However, it should also be recognized that it is premature to conclude that ADM will predict performance in all, or even most, MT environments. ADM has successfully predicted reliable performance measures of dispatching and ATC. But it may be that these particular jobs share specific characteristics not found in other jobs. For example, they both involve the application of limited resources (law enforcement and emergency units in dispatching and air space in ATC). Although LCAC navigation may on the surface seem like ATC, LCAC Navigation does not really involve the application of limited resources. Rather, it involves the figuring of space, time and movement so that a vehicle will arrive at a particular destination at a particular time. Similarly, nurses' decisions do not center on the distribution of limited materials or resources such as medications, equipment, or staff. The central task of the ICU nurse is to integrate many pieces of information about a patient's status and then apply relatively unlimited, at least in most U.S. hospitals, resources to encourage positive changes in the patient's health. The major factor that determines success as a chef is not careful distribution of limited food sources. Rather, successful performance appears to involve the ability to interleave and prioritize tasks to maximize quality and efficiency. Hence, future use of ADM as a commercially viable test that is generally applicable to MT environments first faces the issue of whether or not ADM predicts performance in other MT jobs.

ADDITIONAL ISSUES SURROUNDING ADM AS AN MT ABILITY TEST

There are additional issues surrounding the use of ADM as a predictor that also must be addressed. As can be seen in Table 5, ADM does not assess *ALL* the cognitive operations that our interviewees reported were required by their jobs. Of particular concern is that the cognitive operation of prioritization is less important to performance in ADM than it is to the MT environments studied in this research. A secondary issue is that ADM does not incorporate a planning component. However, we argue that the lack of planning is probably not a factor that would influence the predictive validity of a test based on ADM. In this section, we first discuss the issue of planning.

PLANNING

ADM does not require participants to develop a plan for how they will proceed during conduct of ADM. It is important to remind the reader, here, that by planning we mean the action of preparing a guideline to be used in future execution of tasks that delineates a particular sequence of actions, perhaps with a time component. By this definition of planning, performance of ADM does not require it, although the participant may identify a particular strategy that he or she decides to use after viewing the contents of the bins. It's important to note that most of the existing measures of MT do not include planning components. Three exceptions are the tests designed to assess dysexecutive disorder.

The lack of a planning component may not be critical to the predictive validity of a test that would be based on ADM. One reason that planning may be relatively unimportant is that it is not common to all MT environments. Our analysis of MT environments indicated that planning was an important part of LCAC navigation and operation, Army combat command, and ICU nursing. However, while chefs must prepare their stations so they are ready for orders they will receive on their shift, there is no way to plan their future actions. Because the environment of restaurant food preparation is unpredictable, planning is not possible. Chefs can only react to immediate needs. They cannot be proactive in any meaningful way. While they may make sure they have a sufficient quantity of the appropriate ingredients for whatever is on the menu, they cannot know which of the menu items will be ordered. In contrast, planning is a critical function of operating the LCAC, and the responsibility of the Navigator. That said, the LCAC Navigators we interviewed told us that no mission ever goes according to plan. While it seems, to some Navigators, that plans are made simply to be broken, the true function of the plan is to provide a frame of reference for the mission's events and to encourage anticipation of potential events among the members of the crew. Although not studied in this research, we suspect that planning is not used in other MT jobs, such as dispatching, because of the unpredictable nature of the environment. In fact, one could argue that a test of MT ability should not include a planning component because that would bias it towards relatively predictable environments. The fact that ADM does not include a planning component may actually make it applicable to more MT environments than it would if it did have this feature.

Table 5.
Summary of Cognitive Operations of Existing Measures of MT
(√..... cognitive operation used in Measures)

Cognitive Operations	MET	SET	Greenwich	SYN- WORK	MATB	ADM	Dual Task	Info. Coord.	PRP
Retrospective Memory (STM)	√	√	√	√	√	√	√		
Retrospective Memory (LTM)	√				√				
Prospective Memory	√	√	√	√	√	√			
Monitoring Output of automatic Responses	√	√	√	√		√	√	√	√
WM Updating and Monitoring	√	√	√	√	√	√	√	√	√

Mental Set Switching	√	√	√	√	√	√	√	√	√
Classification	√	√	√	√		√			
Rehearsal for STM Storage	√	√	√	√	√	√	√		
Selective Attention				√	√	√	√	√	
Divided Attention				√	√	√	√	√	
Prioritizing				√	√	√	√		√
Deductive Logic						√			

PRIORITIZING

Perhaps the most important cognitive skill that is not adequately assessed by ADM is the ability to prioritize. In contrast to planning, the need for prioritization is a defining characteristic of MT environments. By definition, MT environments involve more than one task. If more than one task is required, there has to be some way of knowing which are more important than others. The only kind of MT environments that would not require prioritization are those in which it wouldn't matter which task was performed before any other. It is hard to imagine a real-world MT environment where this would be the case. It seems clear to us that prioritization is a critical component of MT environments and it is also clear that ADM does not require prioritization as it is found in real-world MT jobs. Before discussing why we believe the current version of ADM falls short in ecological validity, it is important to understand what that standard should be. Hence, we first discuss how prioritization is demanded by MT jobs.

According to subject matter experts in each of four very different jobs, one of the biggest factors that determine success on the job is the ability to prioritize. If we are to believe our interviewees, lack of a prioritization component is a serious problem for any test of MT ability. In each of the jobs we analyzed prioritization is the ability to identify the one task, of all of the many urgent tasks, that should be accomplished first. Prioritization is so important that participants from each job told us that simple mnemonics are used to remind the worker about priorities. Nurses for example, use an ABC mnemonic to code for *Airway, Breathing, and Circulation*. If a patient has an obstructed airway, there is no sense in checking for circulation. Similarly, if an airway is open, but the patient isn't breathing, the nurse should focus on breathing and, when that is established, turn to circulation. The LCAC Navigators reported that they used an aviation mnemonic to establish that maneuvering the craft (aviate) was the first priority followed by navigate and communicate, in that order. Again, there is no sense in attempting to navigate if there's an obstacle in front of the craft. The craft must first be moving in some direction, which direction is a secondary issue.

The ability to prioritize appears to be an important individual difference factor that determines job success. Each of the participants told us that they had encountered individuals who never learned how to prioritize. We heard many examples of people who were simply overwhelmed by the number of tasks and the complexity of organizing them. For example, the Four-star General we interviewed told us that many "junior" officers, even those who've reached the rank of Lieutenant Colonel or even Colonel, fail to understand that while *every* task is urgent, the completion of some can

be delayed longer than the completion of others. In military combat situations, all tasks truly are urgent. However, the harsh reality is that a leader must prioritize because he cannot simultaneously perform all tasks. To many junior officers, every task has the same priority and they become impatient with a commander who chooses to focus on one task before turning to another.

The ability to effectively prioritize is also important to LCAC Navigation, nursing, and food preparation in restaurant kitchens. One of the ICU nurses told us that he's seen many nurses who miss the big picture, or the purpose, of their job. He told us about a nurse whose main task at the time of the incident was to get his patient released from the hospital. But instead of preparing the patient and his/her belongings, he was busy charting and updating the patient's records, which, of course, could be done after the patient had been released. He also told us about a nurse who delayed giving a patient badly needed pain medication for 40 minutes because he/she was busy with another task and didn't want to be interrupted. One of the chefs we interviewed told us that she could tell within 15 minutes of working with someone new whether they were going to make it in the kitchen. She knew that if they couldn't set appropriate priorities, they'd become overwhelmed or would take much too long to perform the necessary tasks. Finally, one of the LCAC Navigators reported that those who cannot perform the job fail because they get overwhelmed in details. Those who keep it simple, and remember to first aviate are able to perform well. Those that keep trying to navigate when they're lost (don't know their exact position) may collide with the more important obstacle. Being lost is bad, but damaging the craft is worse. If the ability to prioritize is important to successful job performance, and there are significant individual differences seen among working individuals, then it is probably important that a test of MT ability include a requirement to prioritize among tasks.

The ability to prioritize apparently is learned from on-the-job experience. The highly experienced Army officer told us that he learned that lesson by experiencing the consequences of *not* performing truly urgent tasks. Similarly, everyone we talked to told us that they were not good at prioritizing when they were novices. Hence, the ability to prioritize appears to be based on well-learned and in-depth knowledge of complex situations in each domain. The kind and severity of the consequences would vary, of course, across domains such that there is no way to teach prioritization as a general skill. It involves weighing of potential outcomes and the interactions among those outcomes. The ability to prioritize is clearly related to the encoding and retrieval of knowledge stored in LTM. While every normal person has the ability to build and use LTM, some may lack the ability to learn from experience. The negative consequences of poor prioritization that our participant experienced as a junior officer leading a company infantry unit may have taught him well what tasks can be postponed for a few more minutes than others. However, it appears that some people fail to learn that lesson. There is another possible explanation why prioritization seems to be a major factor that determines performance in MT environments. Prioritization is a metacognitive task that can only be undertaken if the basic job performance tasks are not using up all of the available resources. By this explanation, one can engage in prioritization (after one has sufficient practice) when one can conduct basic tasks and still have sufficient cognitive resources to set up and monitor a list of priorities. In other

words, prioritization may be performed only when cognitive resources are available because the basic tasks have been performed efficiently. It is clear prioritization is critical to performance. Whether it is important because of knowledge gained through experience or because it can only be performed when other tasks are performed efficiently is an issue that should be addressed by future research.

It may be possible to measure the ability to learn effective prioritization schemes. However, the test would have to provide appropriate feedback at appropriate times so as to make learning of priorities possible. Subject matter experts in MT environments appear to learn at least two different kinds of prioritization schemes, each based on consequences of successful task completion. First, *structural* relationships among some tasks in some MT environments *demand* that certain tasks be performed before others. In such cases, there is a relationship among the tasks such that successful performance of lower priority tasks depends on the performance of higher priority tasks. For example, nurses know that circulation and breathing depend on the existence of an open air way because one cannot breath without an airway and there is nothing to circulate if oxygen is not taken into the system. Analogously, LCAC operators and Navigators know that they can't go *anywhere* if they run into an obstacle. There is a hierarchical relationship between aviation, navigation, and communication in that the last two depend on the successful performance of the first. The natural world in conjunction with goals (e.g., keep the patient alive or get the LCAC to a particular destination at a particular time) affords and demands these prioritization schemes. The first task *must* be performed before the second, and the second before the third, because it makes no sense, or it is impossible, to perform them in any other sequence. In essence, there is a hierarchical relationship among the tasks that determines the relative priorities they are given. Herein, we will refer to this type of prioritization scheme as *structural*.

Second, prioritization of tasks may be established because (1) the positive consequences of one task may be more valuable, (2) the negative consequences of one task may be less desirable, (3) a particular task must be performed within a narrow temporal window, or (4) timing factors make it more expedient to perform one task before another. For example, when faced with two patients, one who has just vomited and another whose blood pressure is dropping, the ICU nurse will attend first to the patient who will suffer the most severe consequences. Notice that there is no structural relationship or dependency between these two tasks. The nurse could attend to the patient who is vomiting before the patient whose blood pressure is dropping, or vice versa. The priorities that are assigned to the tasks depend on the relative value of the consequences that occur as a result of successful completion of the tasks.

PRIORITIZATION AND THE CURRENT VERSION OF ADM

We now turn to how ADM encourages prioritizing tasks, why we believe it does not adequately assess this important ability, and how it could be simply made to do so. First, consider how, at least on the surface, ADM seems to encourage prioritization. ADM requires the user to perform a set of tasks: (1) classify objects in bins based on bin and object attributes, (2) query the system about the attributes of particular objects, (3) study the attributes of each bin, (4) identify available objects, and (5) refer to object numbers when classifying or querying. The current version of ADM does indeed set up

a hierarchical relationship among the tasks. Ostensibly, participants must attend to items when they become available so as to encode their item numbers. The system requires that the participant make reference to items numbers when classifying or querying items. Hence, one must have some knowledge of available items and their designation. Second, successful classification depends on information obtained from querying objects. One must have some knowledge about the attributes of a particular object to classify it. One must also have knowledge of the bin contents to successfully classify objects. Hence, on the surface, it seems that ADM requires participants to prioritize. However, the need to prioritize can be bypassed in the current version. But before considering how participants can step around the need to prioritize tasks, let's look at one more way in which ADM encourages prioritization.

ADM also uses negative and positive consequences, in the form of points earned, to encourage prioritization. For example, one scores better over all if older objects are dealt with first as the overall score is determined by the average "life" of objects. Any errors a participant makes and any inefficiency in conduct of the tasks add time and reduce the overall score. One also earns more points for classifying items into bins that specify each of the three possible object dimensions: shape, size, and color. For example, ADM awards three points for classifying a red, tiny, circle into its appropriate bin and only 1.45 points for classifying a red object of any size or shape. Hence, ADM incurs positive consequences for complete attribute querying and for quick decision making. These rewards should encourage users to prioritize certain tasks over others. In language we previously used, some tasks should be given high priority because the consequence of their successful completion is more valuable. ADM also appears to demand prioritization because of the ostensible structural and hierarchical relationships among the tasks. In practice, however, the positive consequences and structural relationships among tasks in the current version of ADM may not encourage prioritization.

We have identified three potential issues with ADM that makes its assessment of prioritization questionable. One issue concerns the perceived relative value, and hence priority, of ADM's various tasks' consequences. The second two issues involve the hierarchical relationships among the tasks.

THE ISSUE OF FEEDBACK AND PRIORITIZATION IN ADM

The issue of perceived relative value of the tasks revolves around how ADM provides feedback to participants. Feedback concerning efficiency is not made available to participants during task execution. The only feedback that is given to subjects during conduct of ADM is the points earned for each classification attempt. If the classification has been successful, the system shows points earned. Here, the participant sees that he/she was given 3.0 points or 1.45 points, for example, for assigning an object to a particular bin. If the classification has not been successful, the system simply indicates so and does not show points lost, or extra time taken, or change to the overall score. The only time one sees one's overall score is at the very end of the session. How that score was calculated is not made available to the subject, so it is difficult to relate the overall score feedback to the many decisions that were made during the session. To gain an understanding of how one's actions were related to overall score it would be necessary to complete many sessions of ADM, perhaps varying strategies. Even then, participants

would likely exhibit the kind of erroneous and overly elaborate beliefs found in hypothesis testing studies. While the instructions to the current version of ADM do state that overall score is determined by how quickly objects are classified, the feedback during and immediately after actual performance cannot easily be interpreted. Because learning depends on feedback, it is difficult for ADM participants to learn effective prioritization schemes that are based on scores.

The lack of easily interpreted feedback in ADM may actually increase its ecological validity to real-world MT jobs. In many jobs, the worker is "working in the dark" so to speak, because they do not receive direct feedback about the success of their decisions. Many MT environments lack direct feedback for each decision that is made and that feedback may be delayed in time. Hence, it may be difficult or impossible to relate consequences to decisions or actions. This may be the reason that it takes a long time to achieve expertise in setting priorities, and that some people never learn to effectively prioritize. It would take many experiences to learn from an environment in which many actions over a period of time could only be indirectly related to a consequence, e.g., overall mission was accomplished. However, a test that has the purpose of predicting real-world performance or learning should *not necessarily* provide an isomorphic simulation of the intended environment in which predicted performance will occur. A practical test must necessarily be short, taking no more than 30 minutes. To provide a reasonable assessment of the learning that might occur over many years in real-world MT environments, it may be necessary to make the feedback about prioritization more direct than would be found in real MT situations. The current version of ADM may not provide a reasonable environment in which the participant can learn prioritization. Hence, it may not assess this important ability. On the other hand, tests do not typically include feedback during performance. Scores are usually only obtained after the test has been completed. The very presence of feedback is controversial in a test of ability as it has direct effects on learning.

If feedback is to be included in a test based on ADM, it is possible to modify it such that the participant is given information that can be directly related to the efficiency their current strategy affords. This might enhance the demand ADM makes on the ability to prioritize. For example, participants could be shown points subtracted from the overall points earned for placing an object in a bin because of the "lengthy" time it took for an object to be classified. It is also possible to change the rules of ADM such that classification of certain kinds of objects (e.g., red ones) are explicitly given a higher priority reflected in higher points earned. The instructions of ADM would be changed to reflect the relative priorities and the feedback received for each item classified would also reflect the variations in priorities. These modifications to ADM would be minor in that the task would operate pretty much like it currently does. However, we cannot know the effect the changes would have on participants' performance. Whether or not these modifications to the current version of ADM would assist participants in forming prioritization schemes is, of course, an empirical question. More to the point, whether these modifications would provide a better assessment of the test's ability to assess prioritization is an empirical question. It may be that learning the right prioritization scheme, under the condition where direct feedback is unavailable, is an important component of MT ability, which would argue for using the feedback process of the

current version. Ultimately, the test's value will be in how well it predicts job performance, simulated or actual. One version of ADM that includes direct feedback about efficiency may better predict job performance than the current version, or vice versa. In any case, we believe the issue of feedback is an important test development issue that should be addressed in future research.

THE ISSUE OF STRUCTURAL RELATIONSHIPS AMONG TASKS IN ADM

Although on the surface it appears that ADM establishes a relative structural hierarchy among its tasks, the current version actually permits bypassing the hierarchy. It is possible in ADM, to bypass the task of attending and encoding the number of a newly available item by deducing its numerical designation. The current version of ADM assigns numbers to items in the sequence they become available to the participant. That is, item #15 is made available for querying first, then item #14 is made available, then item #13, and so on counting down in sequence. Because of the systematic numbering of items it is not really necessary to attend to item numbers when they become available. It is not even necessary to attend to the display that indicates any item has become available. If a participant ignores the item number of a newly available item, as long as he/she knows the number of one of the available items, he/she can guess about all other item numbers. In essence, the task of knowing which items are available is not essential. If the participant guesses that item #12, for example, is available and attempts to query the system for that item, the system will respond positively if it is available and negatively if it is not. The allowance to guess number designations reduces WM load and removes one task from the participant's slate. Guessing items numbers is truly inefficient and it will have a negative effect a participant's overall score. However, the lack of feedback regarding inefficiency may make it such that participants are unaware that guessing has any detrimental effect.

But, what if the numbers didn't come up in sequence? The participant would then have to pay close attention to item numbers as they became available because they would have no way of guessing what the item designations were. If a participant had not paid attention to the item number and failed to store it in STM, all other tasks concerning that item could not be performed because the system requires the input of the designation for querying and assigning. At the very best, the participant might be able to guess the number designation of the item, but they would most likely be wrong. Just as establishing an open air way would be the first priority for a nurse, getting the item number right would be the first priority in ADM because all other tasks would depend on it. Of course, requiring participants to pay close attention to item designations may well increase WM load in ADM, which may in turn make the task more difficult than it already is. It could make ADM better at discriminating performance at the upper end of the distribution, or it could produce a floor effect. This is clearly an empirical question about ADM that should be addressed.

It is also possible to bypass full and complete querying of an object in the current version of ADM. For example, suppose the system provides three bins in which to place objects. The first bin takes any object that is red. The second bin takes only red tiny circles. The third bin takes medium squares of any color. One strategy a participant could take is to first query color of an object. Now suppose the system returns the value

of "red" for an object. Even if the object actually was a red tiny circle, the participant could then classify it into the first bin and be correct. In the current version of ADM, the participant would receive fewer points for assigning the object to the first bin, but there would be no punishment if, in fact, the object was a red tiny circle. Here, there is a less stringent relationship between querying and classifying than is possible. While accurate classifying depends on knowledge of *some* attributes, it does not depend on knowledge of *all* attributes. This feature of ADM was intentionally built in. It addresses the issue of optimal stopping where an individual decides to not gather all the possible information, but only the essential information. Initial studies of dispatching revealed that experienced personnel respond to time pressure and emergency by extracting only information that is essential to their most pressing task. Hence, ADM currently allows participants to set a kind of priority, weighing the speed against the thoroughness of classification. Hence, the feature may actually be a strength of ADM. Additional research should consider evaluating the contribution this feature makes to the predictive validity of ADM.

SUMMARY

In summary, we believe a greater emphasis on prioritization within ADM may increase its criterion and ecological validity. As we have discussed, it is possible to modify ADM in any number of ways to greater emphasize the setting and use of priorities. Some modifications address the relative value of tasks in terms of consequences. Other modifications address the issue of structural dependency among tasks. However, *modification should be approached cautiously* because ADM has already been shown to predict simulated job performance. It is entirely possible that modification may only result in lowering its predictive capabilities. At the very least, any modification to ADM must be evaluated to determine how it affects the test's psychometric properties of reliability and validity. On the other hand, examination of the effect of prioritization components on test validity is an important consideration. Both kinds of prioritization schemes (based on structural dependency and relative value) are critical elements of MT environments and more research is needed to determine how important they are to MT ability.

Investigation of the relative effect of varying prioritization requirements is also important to the development of the test. As we discuss in the next section of this report, current professional standards for educational psychological tests call for clear definition of the construct to-be-measured based on scientifically sound investigations of the underlying ability. In truth, very little is known about MT as it occurs in real-world environments. Very little is known about its measurement. While ADM's ability to predict performance in some jobs is extremely encouraging, only four experiments have tested it. Moreover, the results of those four experiments can be found in one article (Joslyn & Hunt, 1998). There is much work to do to lay the groundwork for a reliable and valid test of MT ability.

CHAPTER SIX: DEVELOPMENT OF AN MT ABILITY TEST

In this section we discuss progress made toward the development of a test of MT ability. Although full development of the test is beyond the scope of the current project, the initial phases of design have been completed. To ensure that the proposed test of MT ability meets criteria recognized by the scientific, educational, and testing communities, design was guided by current testing standards published jointly by the American Educational Research Association (AERA), American Psychological Association (APA), and the National Council on Measurement in Education (NCME) (1999). Using the standards to guide the process of test development and evaluation also ensures the MT test (1) will be of the highest quality, (2) can be safely used by government agencies and private industries, and (3) can be commercialized. Finally, the standards provide a framework on which to organize and evaluate the development process.

TEST STANDARDS

Development of an MT ability test was approached with careful consideration of current standards (AERA, APA, NCME, 1999). The AERA et al. (1999) document prescribes and describes a four-phase approach to test development and provides enumerated criteria that all educational and psychological tests must meet.

According to the standards, the test development process begins with a statement of purpose and the construct or content domain to be measured. The second phase involves establishing test specifications such as the number of items, response formats, and time restrictions. These specifications usually form the basis for later test evaluation. If the finished test meets the specifications, it is positively evaluated. In the third phase of test development, test items are compiled, instructions are developed and a draft test is fielded with the purpose of evaluating the test items. The test is then revised and submitted to the fourth phase of test development, which involves its evaluation for operational use. These four development phases prescribed by current standards are outlined below, showing the main points to be considered in each phase.

- 1) Phase I. Delineation of the purpose of the test and scope of the construct
 - Extend original statement of purpose and construct into a framework that describes the extent or scope of the construct
 - Delineate aspects (content, skills, processes, and diagnostic features of the construct to be measured)
 - Develop framework guided by theory and/or analysis
 - Use framework to guide subsequent test evaluation
- 2) Phase II. Development and evaluation of the test specifications
 - Delineate format of items, tasks, questions,
 - Delineate response format or conditions
 - Delineate type of scoring
 - Set desired psychometric properties such as difficulty and discrimination

- Indicate desired test difficulty, inter-item correlations, and reliability
 - Specify time restrictions
 - Specify characteristics of intended population of test takers
 - Specify procedures for administration
 - Establish normative and/or criterion references
- 3) Phase III. Development of field testing, evaluation and selection of items and scoring guides and procedures
- Assemble items into a test
- 4) Phase IV. Assembly and evaluation of the test for operational use

Appendix B summarizes the six sets of standards given in the AERA et al. (1999) document pertinent to test development. These standards are relevant to a test's (1) validity, (2) reliability, (3) development and revision, (4) scales and norms, (5) administration, and (6) documentation. For a variety of reasons, some of the standards given in Appendix B are not applicable to the MT ability test envisioned in this research. For example, some standards are specific to tests that employ extended response formats such as essay tests, which will not be a component of the MT ability test designed thus far. Those standards that are not applicable to the present test development effort are shown in gray text in Appendix B. Important standards that must be considered are shown in black text.

OVERVIEW OF MT TEST DEVELOPMENT

The MT test will be based on Joslyn's and Hunt's (1998) ADM task for reasons previously discussed. Full development of a test that would meet current standards (AERA, APA, NCME, 1999), however, requires additional research. Previous research and the present study provide a sufficient understanding of MT, as an ability and psychological construct, to specify the purpose and scope of the test. A framework for the test can be developed at this point, which should describe the extent of the domain to be assessed and the scope of the construct (AERA, APA, NCME, 1999). The framework should also specify aspects of the construct to be measured such as the content, skills, processes, and diagnostic features. The present research has provided the necessary analysis and understanding of MT environments and measures on which a test framework can be based. A description of the purpose, scope, and framework of the test are provided later in this section of the report.

Additional research, however, is necessary to complete the second phase of test development, which requires test design to be taken to a higher level of specification. For example, additional research is needed to determine whether an acceptable level of criterion validity could be obtained with two sessions of ADM, vs. the four sessions investigated in Joslyn's and Hunt's (1998) original studies. At this point, because very few studies have researched ADM, it is not possible to determine how many sessions

(or "trials" in the parlance used in the standards) would provide a stable measure of MT ability.

There are also important questions about the response format used in ADM that need to be addressed before specifications for a test can be developed. ADM currently requires keyboard responses. The participant must type in an "o", "a", or "b" to query an object, assign an object, or look at the bin contents. When querying an object, the participant must type in "size", "shape", or "color" to obtain information about these dimensions. The system also requires a variety of other responses, all typed in through the keyboard. While typing may be a component of some real-world MT jobs, it certainly is not common to all. The fifth experiment conducted by Joslyn and Hunt (1998) revealed a small, but significant, relationship between individual differences in typing skill and performance differences found in ADM ($r = .33$), which suggests that typing ability may contribute to ADM performance. Whether the inclusion of typing as a response format increases or decreases the predictive capability of ADM cannot be determined at this time. Because typing is yet another task of the many that must be performed in ADM, it may help to create a good simulation of an MT environment. On the other hand, it might reduce the scores of those who "hunt and peck" when they are typing, who might otherwise be excellent multi-taskers. If this is true, the adoption of typing as a response format reduces the validity of the test. Decisions about response format and other test features must be made on strong empirical evidence, according to standards. Moreover, if it is to be used for selection and placement purposes, it will be critical that the test has face validity and receives positive evaluations from industries that are likely to use it. Response format is a key feature that may affect those evaluations, making empirical study of response format critical to the test's success.

In summary, specification of many of the test's important attributes requires additional research. However, it is possible to specify some desired test characteristics such as appropriate test taker populations, maximum test duration, and how the test will be referenced (criterion vs. normed). Where it has been possible to do so, we have delineated these specifications, which are described later in this section of the report.

The third and fourth phases of test development require rigorous investigation of the test's reliability and validity. Examination of the test's construct and criterion validity will be important investigations. Establishing the construct validity of the MT test will be critical to its scientific grounding as well as its commercial viability. In truth, it is not yet clear what ADM measures, nor which of its components are critical predictors of job performance. ADM's construct validity has not been adequately demonstrated. It could be that ADM is largely a measure of WM. Our analysis of MT environments, ADM, and other measures of MT certainly indicate that WM components play a significant part in MT performance. Perhaps ADM taps nothing other than individual variability in using WM processes. Although Joslyn and Hunt (1998) showed that one WM measure was not correlated with ADM, the particular WM measure used in their study is limited, tapping only a few of the WM processes proposed by current theoretical accounts. ADM could also be tantamount to fluid intelligence or processing speed. Again, its relationship to intelligence and other constructs has not been clearly established. In the final section of this report we discuss future studies that will be used to establish the test's construct and criterion validity

and other psychometric properties. At this point, however, we turn to describing the products of the first phase of test development.

TEST PURPOSE, SCOPE, AND FRAMEWORK: PHASE I.

As previously discussed, the first phase of test development focuses on establishing clear definitions of the proposed test's purpose and scope. A framework for the test is developed that extends the purpose of the test to describe the construct to be measured. The framework delineates aspects of the construct that are targeted by the test. What follows documents the intended purpose, scope, and framework for a test of MT ability. Standards (AERA, APA, & NCME, 1999) that are relevant to the points made in this section are given in parentheses.

PURPOSE

The MT test will serve a scientific measurement purpose that can be practically used to address applied needs in MT environments. Broadly stated, the purpose of the test will be to measure individual differences, within normal populations, in multi-tasking ability. In so doing, the test can be used to identify those individuals who are likely to perform well in environments or jobs that require high levels of MT ability. The test will incorporate a scoring system that predicts measures of asymptotic performance in real-world MT environments, as well as measures of time required to reach asymptotic levels. Hence, it will be both a test of ultimate performance and a test of skill acquisition. (Standard 3.2)

MT ability is a psychological construct that has received increasing attention in the basic and applied literature (e.g., Burgess, 2000; Burgess, Veitch, de Lacy Costello, & Shallice, 2000; Joslyn & Hunt, 1998; Meyer & Kieras, 1997; Proctor, Wang, & Pick, 1998; Yee, Hunt, and Pellegrino, 1991) (Standard 3.1). Simple stated, the MT construct is the ability to concurrently perform or interleave multiple tasks. MT ability is thought to place heavy demands on several executive control functions, which many theoretical accounts include as part of working memory (Burgess, 2000; Burgess, Veitch, de Lacy Costello, & Shallice, 2000). Despite its probable overlap with the working memory construct, current findings indicate that MT ability is a distinct individual difference variable (Joslyn & Hunt, 1998). Current findings also indicate that it has little to no relationship to other constructs such as processing speed and fluid intelligence (Joslyn & Hunt, 1998). These conclusions, however, warrant further investigation for reasons previously discussed. MT ability also incorporates the ability to prioritize the many tasks that must be performed. A body of research exists that supports the existence of individual differences in the ability to concurrently perform or interleave multiple tasks. Recent research (Joslyn & Hunt, 1998) has succeeded in measuring such differences and predicting performance in real-world environments and jobs that require individuals to use the ability. The test will be based on a recently developed laboratory task of time-pressured decision-making (Joslyn & Hunt, 1998) that has been shown to be highly predictive of simulated emergency dispatching and ATC job performance. (Standard 1.2, 3.2)

SCOPE

The test is intended to discriminate differences in MT ability among normal populations of adults. Although a body of research has associated MT ability with dysexecutive syndrome and a variety of other neuropsychological disorders that involve impairment of executive control functions (Burgess, 1998; Burgess, 2000; Burgess, Veitch, De Lacy Costello, & Shallice, 2000; Shallice & Burgess, 1991; Wilson, Evans, Emslie, Alderman, & Burgess, 1998), the test is not intended as an instrument to diagnose or otherwise measure such disabilities. The test is intended for adult populations who work in real-world MT environments, and should not be used to discriminate differences among children or aged populations. The test is also intended to have limited criterion validity with respect to work environments. It is intended to predict relevant measures of performance in MT environments, but not in stressful, fast paced, nor time-limited environments; however similar these environments may be to MT jobs. (Standard 1.2, 3.2)

FRAMEWORK

The present research provides a logical framework for understanding MT ability and the proposed MT ability test (Standard 3.1). Standards recognize that this framework may change as test development proceeds through the interplay between construct development and test development (AERA, APA, NCME, 1999). However, current analysis supports basing the MT ability test on the cognitive requirements commonly found in real-world MT jobs. Hence, the MT ability test will incorporate cognitive operations that current analysis shows are critical to successful MT performance. The cognitive operations that appear to be critical are STM rehearsal and storage, WM updating, prospective memory, divided attention, selective attention, mental set switching, LTM retrieval, and prioritization.

Analysis of the ADM task reveals that its current version incorporates and requires participants to employ a set of cognitive operations that are a good match to the operations required by MT environments. Short-term, prospective and working memory operations are integral to both ADM. Executive control functions such as mental set switching, selective attention, divided attention, and rehearsal for STM are also required by ADM.

The ability to effectively prioritize multiple tasks appears to be a critical function that workers must perform in MT environments. While the ability to effectively prioritize multiple tasks in the real world is what makes or breaks a worker, however, we currently do not know if ADM can be performed relatively successfully without this skill. However, it may be possible to increase the degree to which ADM measures the ability to prioritize tasks by modifying ADM's structure, scoring system, or rules. The importance of prioritization to real-world performance in MT jobs warrants investigation of modifications to ADM to better represent the ability to effectively perform this operation.

ADM also fails to incorporate a LTM retrieval component in the sense that domain-specific declarative or procedural knowledge that is typically learned through extensive on-the-job experience is not utilized in ADM. However, any abstract test that would be

applicable to many job domains would *necessarily not* include LTM retrieval in the way it is used in real-world environments. The requirement that the test be applicable to a wide variety of jobs appears to preclude any meaningful LTM retrieval component. Hence, current and modified versions of the ADM task will be designed to measure eight critical cognitive components required by MT environments, which include STM rehearsal and storage, WM updating, prospective memory, divided attention, selective attention, mental set switching, and prioritization. (Standards 1.2, 3.2)

TEST SPECIFICATIONS: PHASE II. (STANDARD 3.3)

While additional research is needed to provide full specification of the proposed test, some of the characteristics the test should possess can be stated. The set of test characteristics that are considered in this section correspond to those prescribed by standards (AERA, APA & NCME, 1999). The specifications that have been fully developed are described in this section, along with those that require further investigation.

Before turning to the specifications, it may help the reader to clarify the use of the term "item", which is used throughout the standards. Many of the specifications prescribed by standards are based on items, as most tests comprise a compilation of such. However, in ADM, and by inference in the proposed MT test, "items" have a different meaning than in most tests. ADM does not include a set of questions to be answered or discrete trials, as most tests incorporate. Numbered questions, in text or figural form, traditionally form the basis for scoring where a total score is calculated on the number of items correctly answered. By analogy, then, items in ADM would refer to the objects that a participant attempts to assign because the current scoring system is based on the number of items successfully classified. Hence, heretofore we interpret "items" in the standards as objects to be assigned to bins.

TEST TAKER POPULATIONS

The test will be appropriate for adults who are otherwise qualified to work in MT environments. These environments may include nursing environments, commercial food preparation in kitchens, emergency dispatching, emergency call receiving, ATC, LCAC navigation, and military combat command, among a host of others. See Table 2 for a list of possible test taker populations. (Standards 1.2, 3.3)

CONTENT AND DIFFICULTY OF TEST: DISCRIMINATION. (STANDARD 1.6)

Generally speaking, the test will employ the current content of ADM as its base. Hence, the number and kinds of tasks required in ADM will also be required in the MT ability test. Using a level of task description that seems appropriate to ADM, the user must interleave the following tasks: (1) querying an object's size, shape or color attributes, (2) assigning an object to a bin, (3) studying bin content, (4) encoding a newly available object designation, and (5) referencing an object for assignment or query. These tasks will form the basis of the test of MT ability.

Also maintained from ADM will be the flow and timing of information displayed on the screen and text-based presentation of all information. Modification of flow and

timing is likely to severely and negatively impact ADM's predictive validity as the relationship between number of tasks and available time is a defining feature of MT environments. ADM appears to provide the right correspondence between tasks and available time to make it a good predictor of other measures of performance. It would be possible and interesting to develop a version of ADM that was either based on the presentation of auditory information or on figural, as opposed to text-based, information. This might allow examination of its relationship to WM constructs such as the visuo-spatial sketchpad and the auditory loop. However, these changes would drastically change what appears to already be working quite well.

While the content of ADM will largely be maintained in the MT ability test, several of the testing standards concerning item difficulty, scoring, feedback and issues of construct validity suggest that additional research be performed so as to ground some decisions about test content in empirical findings.

One example of the need for further research is in the determination of the bins' content, which in part determines test difficulty. When the attributes that define the contents of the bins do not overlap, the task of assigning objects becomes very simple. One only needs to query an attribute, any attribute, and an object can be assigned. When the bins' attributes start to overlap, however, more than one attribute of an object must be queried. Bin content sets up a logical structure that dictates the most efficient means by which to query an object. The greater degree of overlap, the more querying is required, which makes the task generally more difficult. In essence, greater overlap among bin content is tantamount to a greater number of tasks that must be completed. It will be important to determine the level of difficulty that maximizes the test's predictive validity. Difficulty should be set such that the distribution of overall scores among the population of test takers is maximized in range. A test that is too difficult will tend to produce a floor effect or a positively excessively skewed distribution. A test that is too easy will produce a ceiling effect or a negatively skewed distribution. The current version of ADM, which enjoys an impressive level of criterion validity, uses bin contents that overlap to a small degree. However, the effect of bin content on task difficulty and test criterion validity should be examined so as to base decisions about content on firm empirical grounds.

Additional research is also needed to determine the content of the rules, how subjects are instructed about the rules, and how performance feedback is given to subjects. (Standard 2.8) We have discussed the need for a greater emphasis on the requirement and measurement of prioritization in ADM. One way to increase the need for prioritizing is to change the scoring rules. This modification would fundamentally change the content of the test, in terms of instructions, scoring, and feedback. Another way to greater emphasize the setting of priorities is to create structural dependencies among tasks, e.g., change the current number designation of available objects. These potential modifications improve many aspects of ADM. However, they also change the content of the test. Hence, additional research is necessary to establish which version better predicts measures of job performance in MT environments.

Test difficulty is also determined by the number of objects available to be classified. The testing sessions of ADM currently present a new object 50% of the time every 15

seconds. This timing factor may be modified to increase or decrease overall test difficulty to accommodate modifications to bin content, for example. The test must be sufficiently difficult to produce a sufficiently wide distribution of scores to afford high levels of reliability and validity.

ITEM FORMATS

Item format typically refers to how a test item is presented. Although it is possible to modify the presentation of objects in ADM, we see no theoretically driven or practical reason to do so. The arrival of newly available objects is currently announced by a text message displayed on the screen. The attributes of the item are also delivered to the participant in text messages. In response to a color query, for example, the system might return "red". The test of MT ability will maintain the item format used in ADM.

PSYCHOMETRIC PROPERTIES OF TEST AND ITEMS

If we consider "items" to be objects in the current version of ADM, consideration of the psychometric properties of items refers to the relative contribution each correct classification of an object makes to measures of the test's reliability and validity. The test developer typically uses statistics derived from item analysis procedures to examine the goodness of each item. Each item is evaluated according to the statistics and either accepted or rejected. In ADM, however it makes little sense to evaluate individual objects that are to be placed in bins. Instead, the issue that should be addressed is the criteria for the psychometric properties of the bins' contents, because that is what determines how each object should be classified and scored.

Joslyn and Hunt (1998) did not report evaluation of variation in bin content and the resulting effect it might have on ADM's reliability or ability to predict simulated job performance in dispatching or ATC. This is a necessary part of test development, however. The tests predictive validity should be evaluated based on the selection ratio, and other practical factors, it produces when applied to populations for which it is intended. Tests of the current version of ADM showed that it predicts simulated dispatching performance at a high level, $r = .70$. Hence, a target validity coefficient for the test should be approximately .70. However, small differences in predictive validity may be inconsequential. For example, a 30-minute test with high face validity and a validity coefficient of $r = .62$ might have greater utility than a 60-minute test with low face validity and a predictive validity coefficient of .72.

Joslyn and Hunt (1998) do not report reliability estimates for ADM. However, logic and statistics dictate that they must be equal to or greater than the validity estimates. As a rule-of-thumb the validity coefficient cannot, on average, be greater than the square of the reliability, putting the reliability estimate of the ADM task at about .84 or greater. A target for reliability for the MT ability test would then be roughly .80 or above. Internal consistency within sessions and test-retest approaches will be used to estimate reliability. The internal consistency measures will permit examination of performance that occurs without practice. The test-retest reliability estimate must be used to ensure that the rate of work measure is stable (Standard 2.9). Inter-session reliability estimates will also be computed.

ITEM ARRANGEMENT

Item arrangement typically refers to how items are sequenced or presented to the test taker. The proposed test will employ the methods currently used by ADM, which display each newly available object in text form with a number designation on the computer screen.

NUMBER OF ITEMS

The number of items/objects in the current version of ADM is irrelevant to the design of the test as more objects are available for classification than could be performed within each session. See *Time for Testing* for relevant specification concerning test length.

TIME FOR TESTING

The test should take as little time to complete as is possible while maintaining psychometric standards of reliability and validity. The current and tested version of ADM includes two practice sessions and four testing sessions, each five minutes long. Adding time for instructions, questions, and answers the ADM task currently takes about one hour or less to complete. However, it may be possible to reduce the number of practice and/or testing sessions and still obtain a stable measure of performance. Practice effects that occur in ADM have not been studied. We do not know what happens, for example, to performance as participants work from the first to the last testing session. Hence, it is possible that performance becomes asymptotic early, perhaps in the first session. Or the opposite may be true, as the course of skill acquisition on ADM is unknown at this point. This is a matter of empirical study that should be incorporated into the test development process to be performed in the Phase II research.

That said, the MT test's maximum duration may be specified on practical grounds. A test that will be used for placement or selection must be relatively short or it won't be used. Industries that are likely to use an MT test for selection purposes are also likely to require that applicants take other tests. Hence, the MT ability test would likely be one of a battery of tests, suggesting that it should be short. We estimate that the maximum time for testing should be 40 minutes. This could be achieved by reducing the current test to one 5-minute practice session and three 5-minute testing sessions. Empirical assessment of practice over sessions may even suggest that a reliable measure may be obtained in much less time. Reliability should not be sacrificed for convenience. Test duration will be as brief as reliability and validity criteria permit.

DIRECTIONS TO TEST TAKERS

Instructions that precede the test will be clearly written and understandable to 90% of test takers drawn from targeted populations. Instructions used in the current version of ADM will form the basis of revised instructions to be used in the test of MT ability. The content of the revised instructions will reflect any modifications made to ADM. (Standard 2.8)

PROCEDURES FOR TEST ADMINISTRATION (STANDARD 3.6).

The test will be made available on and delivered to subjects via the web. The test will be self-administered in that the test takers will access the appropriate web site, receive the instructions therein, and conduct the practice and testing portions of the test. To be successful, self-administration methods must produce similar distributions of performance in populations tested by Joslyn and Hunt, (1998). (Standard 2.8)

Entry into the web site, and proctoring of test administration, will be overseen by the test developers during the development phase and by the organizations and agencies using the test to assess MT ability in populations of interest. To prevent potential test takers from practicing any part of the test, entry into the test site will be limited to those who have been given passwords that can be used only once.

PROCEDURES FOR SCORING (STANDARD 3.6)

Scoring in ADM is based on the time it takes to classify objects and the number of points earned for each classification. These objective measures are directly computed by the testing program. The procedures for scoring performance on the MT ability test will be those used by the current version of ADM.

RESPONSE FORMAT (STANDARDS 1.7, 3.6)

The response format will be used that meets the following dual criteria of face and predictive validity. The response format will be selected that (1) is rated highest by test takers, and (2) provides the most reliable and predictive measure of MT ability.

NORM REFERENCED OR CRITERION REFERENCED

MT environments most likely vary in the level of MT ability required to perform well on the job. Hence, it is not possible to determine a general criterion performance level that would suit all organizations and agencies that might use the test. For this reason, the test will be able to indicate performance relative to relevant populations. The MT test of ability will be norm referenced.

ASSEMBLY, FIELD TESTING, AND EVALUATION: PHASES III AND IV

Phases III and IV of test development will be conducted in the second phase of this research.

CHAPTER SEVEN: DEVELOPMENT AND VALIDATION OF MT TEST

Many questions might be asked about MT ability and its measurement with ADM. For example, could a figure-based version of ADM be created that would still predict performance in MT jobs? What is it about ADM that makes it such an impressive predictor? Can the response format be changed without affecting criterion validity? What is MT ability and does ADM truly measure it? Science is just beginning to investigate the ability used in real-world MT environments as an individual difference construct. Hence, there is much research that is needed to better understand this apparently important variable.

The purpose of the present research has been to develop a reliable measure of MT ability that can be used to identify individuals who are likely to perform well in MT jobs. This goal serves to guide and limit selection of research hypotheses to be tested. Rather than pursuing questions based on whim, interest, or preference, test standards dictate those that ensure the resulting test is reliable, meets its purpose, and can be safely used for practical purposes.

In this section of the report, research questions that are particularly important to development of a test of MT ability are identified. We then describe a set of studies that would help to resolve issues surrounding (1) test design and assembly, (2) development of MT as a construct, and (3) validation of an MT ability test. Each of these three sets of issues refers to requirements that are incorporated in the third and fourth phases of test development as prescribed by the AERA et al. (1999) standards.

ISSUES OF TEST DESIGN AND ASSEMBLY

The test specifications prescribed by the AERA et al. (1999) standards were given in the previous section. Here, we briefly discuss test specification issues whose resolution must be resolved through additional research.

TEST LENGTH

The test will be only as long as it needs to be to obtain a stable measure of MT ability. Decisions about test length are fed by conflicting motivations in that longer tests produce more reliable measures but practical considerations suggest a short test is needed. But, how long does the test need to be to be reliable? How many sessions of ADM are needed to provide a stable measure? What happens to performance with practice over sessions? Where does performance asymptote? It is also important to establish how, or if, MT ability changes as a function of practice. Standards require evidence concerning the effects of practice and coaching if the test is thought to measure skills or abilities that are not affected by such instructions (Standard 1.9)

The answers to these questions can be answered by looking at performance measures by session. This would permit identification of asymptotic performance, if it occurs in four sessions. The relationship between session performance and criterion measures may be examined to determine when a measure meets criterion validity criteria.

RESPONSE FORMAT

Selecting a response format that is acceptable to the user population may have a substantial effect on the test's success. The perception that the test inappropriately demands too much typing, for example, may negatively impact its face validity and use. From a standards and scientific perspective, the response format may affect the test's ability to predict other measures of performance. The response format that best meets face and criterion validity requirements will be adopted. But what kind of response format is that? Several modifications could be made to the response format currently used by ADM. The requisite typing, for example, could be reduced to single keystrokes. For example, instead of typing in "shape" the test taker could hit "1" on the keypad. Or,

responses could be made by "clicking" on a button using the computer mouse, which has the advantage of being standard for web-based applications. Changes should be approached with caution, however, because different response formats could negatively affect ADM's demonstrated predictive validity.

TEST DIFFICULTY

Various factors are likely to affect test difficulty. We have discussed several issues including variations in bin content and overlap, the interval between item presentations, relationship of feedback to performance, and feedback content and clarity. The test should be difficult enough to discriminate performance at the upper ends of the distribution, but not so difficult as to skew the distribution. It is clear that the distribution should produce a broad range of scores. It is not clear how the various factors that affect difficulty should be set so as to ensure that range.

FEEDBACK

ADM currently provides feedback to the participant whenever an object is assigned to a bin, and at the end of every session. The number of points earned by correctly placing an object in a bin is displayed immediately after the assignment. At this time, the system also displays "Good Match! There are now N objects in this box. You just earned X points." A total score corresponding to total points earned for the session is displayed when the five-minute session has been completed. At that time, average time and game time are also displayed.

As previously discussed, feedback is not usually provided in tests. Feedback such as currently provided in ADM is likely to influence performance and direct participants' strategies. Moreover, only average time is used as a predictor of MT performance in simulated real-world settings. The number of points earned was not used as a predictor in Joslyn's and Hunt's studies (1998). Although the instructions state that performance is a function of how quickly and accurately objects are classified, the current feedback may mislead test takers to focus on points earned rather than on classification speed.

On the other hand, the feedback that is currently provided in ADM may motivate and focus the participant. It may be partly responsible for the high predictive validity demonstrated with ADM. Issues concerning feedback need to be addressed to ensure that the resulting test is fair to all participants, regarded as reasonable by users, and meets stringent psychometric standards. Future research should address issues concerning the amount and kind of feedback provided by the test.

INSTRUCTIONS AND TEST ADMINISTRATION (STANDARD 2.8)

To ensure that the test meets the specifications given in this report, it will be necessary to evaluate the adequacy of the instructions and test administration procedures. It will be necessary to modify the instructions to accommodate presentation on the web and any other changes made that are relevant to the instructions. The test administration procedure will be modified from that used by Joslyn and Hunt, which in some cases involved individual instruction. To maximize ease of use and accessibility, the test will be self-administered and be provided on the web. Although instructions, practice, and test performance will be self-administered, it will probably be necessary to

proctor the test. Any of these modifications could negatively affect the test's psychometric properties. The effects of these changes must be evaluated to determine if they enhance the test, meet current standards of testing, and meet specifications.

Modifications will also be made to the ADM task so as to give it a commercial look and feel. In its present form, it is very clear that the ADM task is a laboratory task, which is not suitable for commercial purposes. It will be necessary to determine if these changes have resulted in diminishing the task's psychometric properties.

EMPHASIS ON PRIORITIZATION

We have argued that the importance of prioritizing in real world MT environments warrants modifying ADM so that it places a greater emphasis on this cognitive element. However, the effects of any changes to ADM, large or small, must be examined to determine if they are beneficial to the overall purpose of the test, or not. Joslyn's and Hunt's findings using the ADM task should serve as a basis for comparison. Any changes made to the task necessitate demonstration that the new version has at least the same predictive power as the original. Hence, once the test is assembled, it will be necessary to replicate some of the Joslyn's and Hunt's original studies. (Standard 1.8)

EMPHASIS ON DEDUCTIVE LOGIC

We have also argued that the role of deductive logic in MT environments is unclear at this point in time. It seems to be a critical component of some jobs, but not so in others. If environments vary substantially in this requirement, should a test of MT ability include a deductive logic component? Moreover, should the construct of MT ability incorporate deductive logic?

The current version of ADM encourages the participant to use deductive logic in evaluating the object attributes described by each bin. We posit that performance in ADM is enhanced when participants deduce the best querying strategy based on the overlap of attributes among the bins. The individual differences produced by the inclusion of this deductive logic requirement in a test are likely to be substantial. These individual differences may positively influence ADM's predictive capability for the jobs examined by Joslyn and Hunt (1998), which include emergency dispatching, ATC, and emergency call receiving. Perhaps these jobs also incorporate a substantial deductive logic requirement that other jobs do not. The incorporation of deductive logic in a test of MT ability should be examined, focusing on whether it increases or decreases the test's ability to predict performance (simulated or actual) in real-world jobs. Also of concern is the effect deductive logic has on the construct validity of MT ability as measured by an MT test.

ISSUES CONCERNING MT AS A PSYCHOLOGICAL CONSTRUCT (STANDARD 1.8)

As previously discussed, it is not yet known whether MT is a separable ability from other psychological constructs such as WM, processing speed, or fluid intelligence. Studies of patients with neuropsychological disorders such as dysexecutive syndrome suggest that the ability to organize and prioritize multiple tasks is orthogonal to intelligence. Patients who otherwise score well on intelligence tests fail at other MT tests

such as the MET, SET and Greenwich tests. Joslyn and Hunt (1998) found a relatively low correlation between one measure of intelligence and performance on ADM. However, the relationship of fluid intelligence to MT ability in normal adult populations has not been adequately investigated. Similarly, the influence of processing speed on MT ability has not been examined in normal adult populations. Finally, theories of WM and MT ability share certain constructs. For example, executive functions of WM that serve to guide attentional resources must also be used in MT situations. It is surprising that Joslyn and Hunt found only a small correlation between one measure of WM, which may not be the best measure, and ADM. These findings beg the question: What is being measured by ADM? The relationship between WM and MT ability warrants further investigation.

ISSUES CONCERNING PSYCHOMETRIC ATTRIBUTES OF TEST

Once the test is designed and assembled, questions about its psychometric properties become the focus of test development. The test's psychometric properties are paramount to its utility. Testing standards dictate that test developers demonstrate that the test produces reliable scores that measure the intended construct. Of central concern are the test's reliability and validity. However, the distribution of scores produced by the test is also important because it largely determines the test's psychometric properties. How does performance vary on the test? Is the test so difficult as to create a floor effect, or too easy such that the scores are negatively skewed? What is the average score? These issues are important to the interpretation of statistics that are derived from the scores. They are also important to interpreting the meaning of test scores, whether the test be norm or criterion referenced. (Standard 2.2)

RELIABILITY

The reliability of the test must first be demonstrated. This issue pertains to the amount of measurement error produced by the scores used in interpreting test performance. Does the test meet standards of reliability given in the stated specifications? (Standard 2.1, 2.3, 2.4, 2.5, 2.6)

CRITERION VALIDITY

If the MT ability test's reliability can be demonstrated, then the next question that must be addressed is its ability to predict a variety of MT jobs. (Standards 1.13, 1.14, 1.16, 1.17, 1.18)

NORMS

What should be considered good performance on the completed MT ability test? What score indicates poor MT ability? To be of use, test scores must be interpreted. Given that the test will be norm referenced, it will be important to establish on an appropriate population the distribution of scores the test produces. (Standards 1.1, 1.2, 4.2, 4.4, 4.5, 4.6)

SEVEN STUDIES TO DEVELOP AND VALIDATE TEST

The issues discussed above are all clearly important to test development and/or validation. Even more clear is that they cannot be resolved by a single study. A hierarchical relationship is evident among the issues. Questions most pertinent to test development (test length, response format, test difficulty, instructions, administration, cognitive components, feedback) must be addressed before psychometric properties (reliability, construct validity, and predictive validity) may be estimated. However, some of the test development issues (e.g., test length and difficulty) must also be revisited once a version of the test has been developed.

To meet the high standards set by the research and testing communities, seven studies have been designed to address the issues we have discussed. The remainder of this report describes these studies.

STUDY #1: TEST ADMINISTRATION PROCEDURES AND INSTRUCTION

Purpose. The primary purpose of the first study will be to assess the effects of changes to test administration procedures and instructions. Changes to the administration procedures will include web-based delivery and self-administration. The instructions will also be changed to adhere to testing standards, make them compatible with self-administration procedures, and develop a commercial "look and feel" to the test. The central question of this study is whether these changes affect performance on the test and the test's ability to predict other measures of MT performance. A secondary purpose of the study will be to examine performance changes over and among the two practice sessions and the four test sessions. This information will be used to make initial estimates of the appropriate test length. A third purpose will be to examine test difficulty. Examination of the score distributions produced by the test will reveal skewness, ceiling, or floor effects. This information can be used to determine if changes in test difficulty should be considered.

Population. This study should recruit participants from a population and sample similar, if not identical, to those used in Joslyn's and Hunt's original studies. In their first study, they used the participant pool at a university. In other studies, college student participants were either recruited through campus wide advertisements in the school newspaper of a university or a community college. In one study, a small population of dispatchers was recruited. The community college population has the advantage of best representing the community at large, most likely having a wider range of abilities and general intelligence levels than found at a university. However, there was no apparent restriction of range problems with the university students used in their first study as some of the highest validity coefficients were obtained with this population. Hence, any of these populations would satisfy the needs of Study #1.

In several of Joslyn's and Hunt's (1998) studies approximately 50 participants were recruited. With attrition, statistics were based on the data produced by slightly less than 50 individuals. Study #1 should recruit at least 50 participants to replicate Joslyn's and Hunt's original studies.

Materials. A new version of ADM will be developed that incorporates web-based self-administration procedures and instructions. To determine if the changes to ADM

have a detrimental affect on the range of scores produced or on the ability of the task to predict other measures of performance, participants will also participate Joslyn's and Hunt's (1998) simulation of 911 dispatching.

Procedure. The strategy of replicating some of Joslyn's and Hunt's original studies using the same measures they used is appropriate for Study #1. The critical issue will be whether surface changes to ADM to make it a commercial polished test have detrimental affects on its ability to predict other measures of performance. Hence, the procedures used by Joslyn and Hunt will be followed to the degree possible.

Results. Below we discuss the data to be gathered and the analyses to be conducted.

Data to be gathered

1. Scores for each subject based on original ADM scoring algorithm.
2. Measures of performance on dispatching test to be correlated with measures of performance on modified test.
3. Individual participant and average performance for each of 2 practice sessions and 4 test sessions.

Analyses to be conducted

1. Descriptive statistics on distribution produced by modified test including measures of central tendency and dispersion. These data can then be compared to analogous statistics reported in Joslyn's and Hunt's original study.
2. Determine if distributions are different.
3. Correlation coefficients with dispatcher simulation
4. Correlation and plots of performance by session to determine how performance changes with practice?
5. Plots and correlation for each session relating test performance to dispatcher task performance

Discussion. If the distribution produced by the original ADM (as indicated by descriptive statistics) are different from the distribution produced by the modified "commercial" version, additional modification to the test may be required depending on how the distributions differ. The most important factor in deciding what to do with the modified version will be its ability to predict dispatcher simulation. If it does not predict as well as the original, the source of the difference would have to be investigated. If the modified version predicts about as well as the original (which must be determined statistically and judged on a qualitative basis) the research focus could turn to the following studies.

STUDY #2: RESPONSE FORMAT

Purpose. The primary purpose of the second study will be to examine issues of response format. The following question is central to this study's purpose. Does changing the response format change the ADM's ability to predict performance in

simulated or real MT environments? A secondary issue will be to determine if changes in response format modify the development of performance with practice.

Population. This study is essentially an extension of Joslyn's and Hunt's original studies. For that reason, many of the design issues are addressed by replicating the methods they have previously used. Hence, a population similar to the one used in Study #1 should be used. A student population of approximately 100 should be recruited. Half of the sample will participate in the modified version of the MT test created in the first study described previously. The other half will participate in a version of ADM that requires a different response format.

Materials. Assuming the web-based version of the new MT test has been positively evaluated by Study #1, it will be incorporated into this study. An additional version of the MT test will be created with a new response format that eliminates the typing required by the old version. The dispatcher simulation will also be employed as a bench marker for criterion validity.

Procedure. This study will again replicate many of the procedures used by Joslyn and Hunt (1998). The critical issue of concern is whether variation in response format alters ADM's ability to predict other measures of MT performance.

Results. Below we discuss the data to be gathered and the analyses to be conducted.

Data to be obtained

1. Scores for each subject based on original ADM scoring algorithm.
2. Measures of performance on dispatching test to be correlated with measures of performance on version created in Study #1 and modified version created for this study.
3. Individual participant and average performance for each of 2 practice sessions and 4 test sessions.

Analyses to be conducted

1. Descriptive statistics on distributions produced by modified test and by new version with different response format including measures of central tendency and dispersion. These data can then be compared to analogous statistics reported in Joslyn's and Hunt's original study.
2. Correlation coefficients with dispatcher simulation.
3. Correlation and plots of performance by session—how does performance proceed with practice?
4. Plots and correlation for each session relating test performance to dispatcher task performance.

Discussion. Response format that produces the best predictive capability will be selected for final version of test.

STUDY #3: FEEDBACK STUDY

Purpose. The primary purpose of the of this study will be to examine how changes in the kind and amount of feedback provided in ADM affect its ability to predict performance in simulated or real MT environments.

Population. A population similar to the one used in the first two studies should be used. A student population of at least 150 should be recruited. Fifty of the sample will participate in the modified version of the MT test created in the first study described previously. Another 50 will participate in a version of ADM in which the way feedback is provided to the participant is varied. The final sample of 50 will participate in a version of ADM in which feedback has been removed.

Materials. Assuming the web-based version of the new MT test has been positively evaluated by Study #1 and has been used successfully in Study #2, it will be incorporated into this study. The response format that produced the highest correlations with the simulated dispatcher test in Study #2 will be used in the base line version of ADM in this study. Another version of ADM will be created in which feedback concerning the participant's performance with respect to both speed and accuracy will be presented. A third version of ADM will be created in which no feedback is presented to the participant. The dispatcher simulation will also be employed as a bench marker for criterion validity.

Procedure. This study will again replicate many of the procedures used by Joslyn and Hunt (1998). The critical issue of concern is whether feedback, or the lack of it, affects the predictive capability of ADM. Each participant will complete one of the three versions of ADM noted in the Materials section of this study, and the dispatcher simulation used by Joslyn and Hunt.

Results. The following describes the data to be obtained and the analysis to be conducted in this study.

Data to be obtained

1. Scores for each subject based on original ADM scoring algorithm
2. Measures of performance on dispatching test to be correlated with measures of performance on version created in Study #1 and modified versions created for this study.
3. Individual participant and average performance for each of 2 practice sessions and 4 test sessions

Analyses to be conducted

1. Descriptive statistics on distributions produced by modified test and by new version with different response format including measures of central tendency and dispersion. These data can then be compared to analogous statistics reported in Joslyn's and Hunt's original study.
2. Correlation coefficients with dispatcher simulation
3. Correlation and plots of performance by session

4. Plots and correlation for each session relating test performance to dispatcher task performance

Discussion. The feedback condition that produces the best predictive capability in this study will be selected for final version.

STUDY #4: PRIORITIZATION

Purpose. The primary purpose of the fourth study will be to examine how changes to the structure of ADM to include a greater emphasis on prioritization will affect its ability to predict performance in simulated or real MT environments. Having established the basic features of the MT ability test, in terms of response format and feedback, we will begin to examine issues concerning the cognitive operations the test requires.

Population. A population similar to the one used in the first three studies should be used. A student population of at least 200 should be recruited. Fifty of the sample will participate in the modified version of the MT test created as the result of the first three studies. Another 50 will participate in a version of ADM in which the structural relationships among tasks will be emphasized. Another quarter of the sample will participate in a version of ADM in which the relative value of task completion is varied. The final fifty will participate in a version of ADM in which priorities are set by locking out certain tasks if not completed within a period of time.

Materials. The web-based ADM task with the response format and feedback that predicted the dispatcher task best will be incorporated into this study. Three new versions of the MT test will be created in which the task will be changed to emphasize prioritization in different ways. In the first, priorities based on structural relationships between the tasks will be emphasized by making the item numbers in a random numerical sequence. In the second version, the relative value of classifying some objects will be increase over others to simulate priorities based on value and consequences. In the third, some tasks will be "locked out" if not attended to within a period of time. The dispatcher simulation will also be employed as a bench marker for criterion validity.

Procedure. This study will again replicate many of the procedures used by Joslyn and Hunt (1998). The critical issue of concern is whether variation in structural and value based prioritization affects the predictive validity of ADM. Each participant will complete one of the three versions of ADM noted in the Materials section of this study, and the dispatcher simulation used by Joslyn and Hunt.

Results. The following describes the data to be obtained and the analysis to be conducted in this study.

Data to be obtained

1. Scores for each subject based on original ADM scoring algorithm
2. Measures of performance on dispatching test to be correlated with measures of performance on version created in Study #1 and modified versions created for this study.

3. Individual participant and average performance for each of 2 practice sessions and 4 test sessions

Analyses to be conducted

1. Descriptive statistics on distributions produced by modified test and by new version with different response format including measures of central tendency and dispersion. These data can then be compared to similar statistics reported in Joslyn's and Hunt's original study.
2. Correlation coefficients with dispatcher simulation
3. Correlation and plots of performance by session
4. Plots and correlation for each session relating test performance to dispatcher task performance

Discussion. The version of the test based on prioritization scheme that produces the best predictive capability will be selected for final version.

STUDY #5: DEDUCTIVE LOGIC DEMAND (BIN OVERLAP)

Purpose. The primary purpose of the fifth study will be to examine how changes to ADM's requirement for deductive logic affects its ability to predict performance in simulated or real MT environments.

Population. A population similar to the one used in the first four studies should be used. A student population of at least 150 should be recruited. Fifty of the sample will participate in the version of the test based on the results of the first four studies. Another 50 will participate in a version of ADM in which the requirement of deductive logic is minimized. The remaining 50 will participate in a version of ADM in which the deductive logic requirement is maximized.

Materials. Assuming the web-based version of the new MT test has been positively evaluated by Study #1 and has been used successfully in Study #2, it will be incorporated into this study. Another version of ADM will be created in which the amount of overlap among bins (in terms of the number of attributes they share) is decreased to a minimum amount. This will minimize the influence of deductive logic in ADM. A third version of ADM will be created in which the number of attributes shared by the bins is increased greater than was used in the original version of ADM. The dispatcher simulation will also be employed as a bench marker for criterion validity.

Procedure. This study will again replicate many of the procedures used by Joslyn and Hunt (1998). The critical issue of concern is how the deductive logic requirement changes ADM's ability to predict other measures of MT ability. Each participant will complete one of the three versions of ADM noted in the Materials section of this study, and the dispatcher simulation used by Joslyn and Hunt.

Results. The following describes the data to be obtained and the analysis to be conducted in this study.

Data to be obtained

1. Scores for each subject based on original ADM scoring algorithm
2. Measures of performance on dispatching test to be correlated with measures of performance on version created in Study #1 and modified versions created for this study.
3. Individual participant and average performance for each of 2 practice sessions and 4 test sessions

Analyses to be conducted

1. Descriptive statistics on distributions produced by modified test and by new version with different response format including measures of central tendency and dispersion. These data can then be compared to analogous statistics reported in Joslyn's and Hunt's original study.
2. Correlation coefficients with dispatcher simulation
3. Correlation and plots of performance by session
4. Plots and correlation for each session relating test performance to dispatcher task performance

Discussion. The version of the deductive logic requirement that produces the highest predictive capability will be selected for final version.

STUDY #6: CONSTRUCT VALIDITY

Purpose. The first five studies have been designed to ferret out issues concerned in test development. Study #6 is the first study to be conducted on a completely designed test. The primary purpose of the sixth study will be to examine the test's construct validity. This study will attempt to resolve questions concerning the relationship of MT ability to other constructs. Is MT a separable construct? Alternatively, is MT ability a component of WM, processing speed, or fluid intelligence. Several models will be developed and evaluated using latent variable analysis.

Population. A sample of participants will be recruited from the college student population. Latent variable analyses require a relatively large sample of participants. Approximately 150 participants students will be recruited for this study.

Materials. Participants will be asked to complete an array of tests for this purpose of validation. Figure 2 depicts examples of potential models to be tested. As shown in the figure, the four sessions of the MT test will serve as indicators for MT ability. The construct of WM will be measured using complex memory span tasks, including reading span (RSPAN), operation span (OSPAN), and counting span (CSPAN) tasks. Other studies utilizing latent variable analyses have successfully used complex span measures as indicators of working memory central executive function (i.e., controlled attention) (e.g., Conway et al. 2002; Engle et al., 1999; Miyake et al., 2000). The construct of STM, which has been statistically distinguished from that of WM (Bayliss et al., 2003; Engle et al., 1999), will be measured with simple forward and backward word and digit span tasks. The construct of Processing Speed (PS) will be measured with digit and

letter copying, and pattern and letter comparison tasks. Finally, fluid intelligence will be assessed with Raven's Standard Progressive Matrices and Cattell's Culture Fair tests.

Procedure. Participants will be asked to complete a battery of tests and tasks noted above that ostensibly measure constructs potentially related to MT ability. A variety of models that describe the relationships among MT ability and other constructs will be tested. Figures 2 and 3 are examples of two models that would be examined. Figure 2 shows a model where MT, WM, STM, and PS all are significantly related, but separable, and are each components of gF. Figure 3 depicts a model in which MT is entirely separable from each of the other constructs. Other models, although not shown here, that show a hierarchical relationship between MT and WM will also be tested.

Results

Data to be obtained

1. Individual scores on each measure.

Analyses to be conducted

1. Descriptive statistics on each measure will be derived including measures of central tendency and dispersion. Statistics that indicate the shape of the distribution will also be derived.
2. Reliability estimates of each measure will be computed.
3. First order correlations among the measures will be computed
4. Latent variable analyses (e.g., confirmatory factor analysis and structural equation modeling) will be conducted to test and compare full, alternative, and nested models pertaining to MT in relation to the other four proposed latent factors.

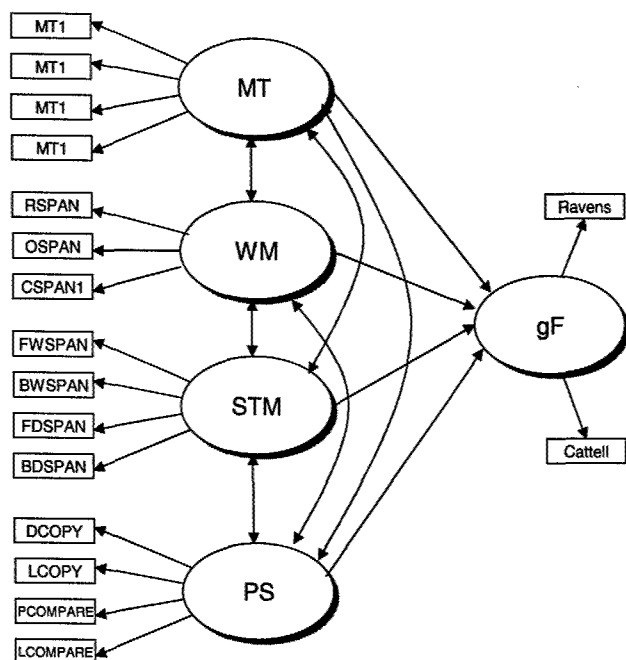


Figure 2. Hierarchical model where MT is a component of gF.

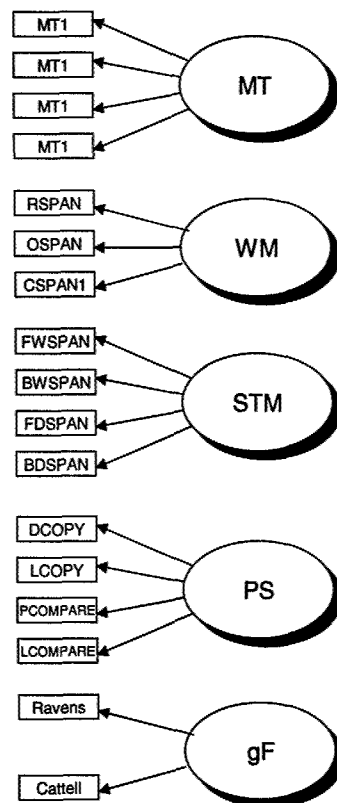


Figure 3. Model where MT is entirely separable from other constructs.

Discussion. The results of this study will be used to evaluate MT ability as a separable construct. They will meet the construct validity requirement of the AERA et al. (1999) standards.

STUDY #7: PSYCHOMETRIC PROPERTIES

Purpose. The primary purpose of the final study will be to examine the psychometric properties of the final version of the test. At this point in time, the research will have produced a completed test. The relationships between the individual differences measured by the new MT ability test and those of other constructs will have been examined in Study #6. It will now be important to establish the degree to which the new version can predict performance in other MT environments. It is important to note that the test development process has, in fact, ensured that the MT test has predictive capability. At every step along the way, the criterion for decisions about test development were based on which version predicted a simulation of 911 dispatching. The consistent use of emergency dispatching simulation provides a necessary stable base of comparison. Attempts to use other measures of performance in other MT environments would only confuse the test development process. However, consistent use of the emergency dispatching simulation also limits the criterion validity of the test. In study #7, this limitation will be evaluated.

The first issue to be examined is the test's stability. Although reliability estimates will have been taken on each of the versions used in previous studies, it will be important to establish the reliability on the final test, which is of course prescribed by standards. The project's previous studies will inform the most appropriate measures of reliability. At this point in time, it appears that test retest and internal consistency measures be estimated.

The second issue to be examined is arguably the most important issue to the test's utility: its criterion validity. How well the test predicts performance in three very different MT environments will be examined. The three environments to be studied are nursing, LCAC navigation, and emergency dispatching. Because the criterion performance measures will vary for each of the three MT environments selected for this study, this final study might be considered three separate studies.

A third purpose will be to provide initial norming data for the final test. If this research is successful, the test will be a practical tool that can be used by the organizations that staff personnel in these three environments. Hence, it will be important to provide norms relevant to each type of job. It would also be useful to provide norms for a general population that represented individuals who might consider career paths in nursing, LCAC navigation, and emergency dispatching. Hence, norm data will also be gathered from community college student populations because this population (1) probably incorporates the broadest range of abilities and general intelligence, and (2) can be readily accessed in numbers sufficient to provide reasonable norms.

Populations. Four populations of participants will be recruited for this study. First, nursing students who are currently serving intern positions at hospitals in their general area will be recruited. Second, future LCAC Navigators who are beginning an LCAC training program will participate. Third, a sample of emergency dispatchers will be recruited. Finally, a large sample of community college students will be recruited for participation. For criterion validity purposes, sufficiently robust statistical analysis requires samples of a minimum of 30 individuals be drawn from the three selected MT environments. A larger sample ($N=200$) will be drawn from the community college student population to adequately represent the distribution.

Materials. All participants will take the web-based version of the new MT test developed from the results of the previous studies. This is the only test the community college students will take. Participants from the nursing, LCAC navigating, and dispatching communities will also be asked to complete tasks in which their jobs are simulated. The dispatching simulation used by Joslyn and Hunt in their studies will be used here. Computer-based simulations will be developed for LCAC navigation and nursing. Additional measures of actual job performance will also be obtained from each of the three communities to determine if the predictive validity observed in previous studies based on simulated job performance generalizes to actual performance in real-world situations.

Procedure. Participants will be asked to complete the MT ability test and the appropriate job simulation task. Measures of actual performance on the job will also be obtained.

Results

Data to be obtained

1. Individual scores on each measure.
2. Measures of actual job performance
3. Individual scores by session on MT ability test

Analyses to be conducted

1. Descriptive statistics on each measure will be derived including measures of central tendency and dispersion. Statistics that indicate the shape of the distribution will also be derived.
2. Reliability estimates of each measure will be computed.
3. Correlation and plots of performance by session to determine effect of practice
4. Plots and correlation for each session relating test performance to simulated job performance measures.
5. First order correlations between MT ability test and relevant criterion measures

Discussion: Utility of Findings. The reported statistics derived from this study will meet standards of scientific evidence and documentation for psychological tests. They will be documented in a test manual.

CHAPTER EIGHT: CONCLUSIONS

The research described in this report broadens and deepens current knowledge of real-world MT. It makes significant contributions to the study of MT. The research provides a way to define MT environments that was previously unavailable to researchers. The definition appears to fit work environments studied in this research and could be used to identify non-MT settings. Future research should evaluate whether the definition provided in this report is useful at discriminating MT settings from environments that do not demand MT.

Comparison of four MT settings and 8 different jobs in those settings showed that although MT environments appear to differ greatly, they share a number of characteristics. However, it may be possible to extend the utility of the characteristics, which currently uses binary classification, by developing a relative grading system in which MT environments are rated along continuous dimensions. For example, the number of tasks required by job could be counted instead of classifying settings as either having many or few tasks. The dynamic nature of an environment might be

graded according to the number of interruptions that occur while other tasks are being performed.

The definition of MT environments has also afforded a path by which cognitive operations that might be demanded by these environments can be specified. In the same way that environmental characteristics might be used to discriminate MT settings from other kinds of work environments, the cognitive operations should be evaluated to determine their utility as a discrimination tool. This research has not provided a comparison between MT environments and other settings, which is needed to better understand the cognitive requirements of each. However, it does provide some tools that could be used in future research.

The cognitive operations have been used in this research to illuminate important aspects of MT. For example, some appear to be more important to MT environments than others. Several appear to characterize complex MT environments from simple ones. Prospective memory, for example, is only important if (1) the worker does not have environmental cues to prompt initiation of a task and (2) the tasks must be interleaved as opposed to being completed serially. Many simple MT laboratory tasks cue each task, or they provide instructions such that prospective memory is not necessary. Prospective memory in real-world MT environments is critical to performance, however. On the other hand, executive monitoring functions that serve to evaluate the outcome of automatic responses do not seem to distinguish or be critical to MT as it is conducted in the applied settings.

This research has also provided a way to identify requirements for a test of MT. A test of MT ability is not yet available to researchers. Because measurement forms the basis of all research, development of a test would greatly advance researchers ability to study MT. Historically, measurement of individual difference constructs is a fruitful endeavor that advances understanding. The present research lays the groundwork for measurement of MT to begin. Initial test design has been completed according to standards and a series of studies necessary to further test development and evaluation have been designed. Future research that addresses the research issues discussed in this report will produce a greater understanding of what is now a very common activity in our world.

REFERENCES

- Ackerman, P.L., Schneider, W., & Wickens, C.(1984). Deciding on the existence of a time-sharing ability: A combined methodological and theoretical approach. *Human Factors*, 26, 71-82.
- Ackerman, P., & Kanfer, R. (1993). Integrating laboratory and field study for improving selection: Development of a battery for predicting air traffic controller success. *Journal of Applied Psychology*, 78, 413-432.
- Allport, A. (1993). Attention and control: Have we been asking the wrong questions? A critical review of twenty-five years. In D.E. Meyer & S. Kornblum (Eds.), *Attention and performance XIV* (pp. 343-350). Cambridge, MA: MIT Press.

- American Educational Research Association, American Psychological Association, National Council on Measurement in Education. (1999). *Standards for Educational and Psychological Testing*. Washington, DC., American Educational Research Association.
- Anderson, J.R. (1993). *Rules of the mind*. Hillsdale, NJ: Erlbaum.
- Anderson, J., Bothell, D., Douglass, S., Haimson, C., & Sohn, M-H. (2002). *CMU-ASP: ACT-R Learns to be an Anti-air warfare coordinator*. Department of the Navy, Office of Naval Research, <http://act-r.psy.cmu.edu/workshops/workshop-2002/talks/>.
- Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. In K. W. Spence, and J. T. Spence (eds.), *The Psychology of Learning and Motivation: Advances in Research and Theory*, Vol. 2. New York: Academic press.
- Baddeley, A. D. (1996). Exploring the central executive. *Quarterly Journal of Experimental Psychology*, 49A, 5-28.
- Baddeley, A. D. (2000). Short-term and working memory. In E. Tulving and F. I. M. Craik, (Eds.), *The Oxford Handbook of Memory*. New York: Oxford University Press.
- Baddeley, A. (2002). Is working memory still working? *European Psychologist*, 7, 85-97.
- Baddeley, A., Chincotta, D., & Adlam, A. (2001). Working memory and the control of action: Evidence from task switching. *Journal of Experimental Psychology: General*, 130, 641-657.
- Baddeley, A., Della Sala, S., Papagno, C., Spinnler, H. (1997) Dualtask performance in dysexecutive and non dysexecutive patients with a frontal lesion. *Neuropsychology* 11, 187-194.
- Baddeley, A.D., & Logie, R.H. (1999) Working memory: The multiple-component model. In Akira Miyake & Priti Shah (Eds). *Models of working memory: Mechanisms of active maintenance and executive control* (pp. 28-61). New York, NY, US: Cambridge University Press.
- Bayliss, D. M., Jarrold, C., Gunn, D. & Baddeley, A. D. (2003). The Complexities of Complex Span: Explaining Individual Differences in Working Memory in Children and Adults. *Journal of Experimental Psychology: General*, 132,
- Braun, R., & Wickens, C.D. (1986). Time-sharing revisited: Test of a componential model for the assessment of individual differences, *Ergonomics*, 29, 1399-1414.
- Broadbent, D. E. (1958). *Perception and Communication*. London: Pergamon Press.
- Brocke, B., Tasche, K. G. & Beauducel, A. (1997). Biopsychological foundations of extraversion: Differential effort reactivity and state control. *Personality and Individual Differences*, 22, 447-458
- Brookings, J.B. (1990). A confirmatory factor analytic study of time sharing performance and cognitive abilities. *Intelligence*, 14, 43-59.
- Brookings, J.B., & Damos, D.L. (1991). Individual differences in multiple-task performance. In D.L. Damos (Ed.) *Multiple Task Performance* (pp. 363-386). Washington, DC: Taylor & Francis.
- Burgess, P. W. (1998). Real-world multitasking from a cognitive neuroscience perspective. In S. Monsell & J. Driver (eds.) *Control of Cognitive Processes: Attention and Performance XVIII*. Cambridge, MA: MIT Press.

- Burgess, P. (2000). Strategy application disorder: The role of the frontal lobes in human multitasking. *Psychological Research*, 63, 279-288.
- Burgess, P.; Veitch, E., de Lacy Costello, A., Shallice, T. (2000). The cognitive and neuroanatomical correlates of multitasking. *Neuropsychologia*, 38, 848-863.
- Caldwell, J. & Ramspott, S. (1998). Effects of task duration on sensitivity to sleep deprivation using the multi-attribute task battery. *Behavior Research Methods, Instruments & Computers*, 30, 651-660. p72
- Carlson, R.A., Sullivan, M.A., & Wenger, J.L. (1993). Coordinating information from perception and working memory. *Journal of Experimental Psychology: Human Perception and Performance*, 19, 531-548.
- Comstock, J. R., & Arnegard, R. J. (1992). *The multi-attribute task batter for human operator workload and strategic behavior research*. (NASA Tech. Mem. No 104174. Hampton, VA: National Aeronautics and Space Administration, Langley Research Center.
- Conte, J., Rizzuto, T., & Steiner, D. (1998). A construct analysis of individual-level polychronicity. *Journal of Managerial Psychology*, 3, 269-287.
- Conway, A., Cowan, N., & Bunting, M. (2001). The cocktail party phenomenon revisited: The importance of working memory capacity. *Psychonomic Bulletin and Review*, 8, 331-335.
- Conway, A., Cowan, N., Bunting, M., Theriault, D., & Minkoff, S. (2002). A latent variable analysis of working memory capacity, short-term memory capacity, processing speed, and general fluid intelligence. *Intelligence*, 30, 163-184.
- D'Esposito, M., Detre, J. A., Alsop, D. C., Shin, R. K., Atlas, S., & Grossman, M. (1995). The neural basis of the central executive system of working memory. *Nature*, 378, 279-281
- De Jong, R. (1995). The role of preparation in overlapping-task performance. *The Quarterly Journal of Experimental Psychology*, 48A, 2-25.
- De la Casa, L. G., Gordillo, J. L., Mejías, L. J., Rengel F., & Romero M. F. (1998). Attentional strategies in Type-A individuals. *Personality & Individual Differences*, 24, 59-69.
- D'Esposito, M., & Postle, B. (2002). The neural basis of working memory storage, rehearsal, and control processes: Evidence from patient and functional magnetic resonance imaging studies. In, Larry R. Squire, & Daniel L. Schacter (Eds), *Neuropsychology of memory* (pp. 215-224). New York: Guilford Press.
- Deutch, J. A., & Deutch, D. (1963). Attention: Some theoretical considerations. *Psychological Review*, 70, 80-90.
- Dickman, S. (1990). Functional and dysfunctional impulsivity: Personality and cognitive correlates. *Personality and Individual Differences*, 58, 95-102.
- Dickman, S. (2000) Impulsivity, arousal, and attention. *Personality and Individual Differences*, 28, 563-581
- Dickman, S.J., & Meyer, D.E. (1988). Impulsivity and speed-accuracy tradeoffs in information processing. *Journal of Personality and Social Psychology*, 54, 274-290.
- Emerson, M., & Miyake, A. (2003), The role of inner speech in task switching: A dual-task investigation. *Journal of Memory and Language*, 48, 148-168.

- Emerson, M.J., Miyake, A., & Rettinger, D.A. (1999). Individual differences in integrating and coordinating multiple sources of information. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 25, 1300-1312.
- Engle, R., Tuholski, W., Laughlin, J., & Conway, A. (1999). Working memory, short-term memory, and general fluid intelligence: A latent variable approach. *Journal of Experimental Psychology: General*, 128, 309-331.
- Eysenck, H. (1997). Personality and experimental psychology: The unification of psychology and the possibility of a paradigm. *Journal of Personality and Social Psychology*, 73, 1224-1237.
- Fischer, S. C. & Harp, S. F. (1997). Incidents in infusion pump use: A nationwide survey of hospital nurses. Santa Barbara, California: Anacapa Sciences, Inc.
- Gopher, D. (1982). A selective attention test as a predictor of success in flight training. *Human Factors*, 24, 173-183.
- Gopher, D., Armony, L., & Greenshpan, Y. (2000). Switching tasks and attention policies. *Journal of Experimental Psychology: General*, 129, 308-339.
- Gopher, D., & Kahneman, D. (1971). Individual differences in attention and the prediction of flight criteria. *Perceptual and Motor Skills*, 33, 1335-1342.
- Harris, W. C., Parasuraman, R., Zinni, M., Hancock, P. A., Harris, S. C. (2002). Viking: Development of an Infantry Scenario Multi-task Environment. Paper presented at APA Midyear Symposium.
- Hegarty, M., Shaw, P., & Miyake A. (2000). Constraints on using the dual-task methodology to specify the degree of central executive involvement in cognitive tasks. *Memory & Cognition*, 28 (3), 376-385.
- Hunt, E. (1986). Information processing approaches to attention. In S.E. Newstead, S.H. Irvine, & P.L. Dann (Eds.) *Human assessment: Cognition and motivation* (pp. 27-42). Dordrecht, The Netherlands: Martinus Nijhoff Publishers.
- Hunt, E., & Pellegrino, J.W. (1985). Using interactive computing to expand intelligence testing: A critique and prospectus. *Intelligence*, 9, 207-236.
- Hunt, P., Linnville, S., Stuster, J., Schneider, K., & Braun, D. (1993). The development of permanent medical standards for landing craft air cushion (LCAC) crew personnel. (Report #93-26). San Diego, Ca, Naval Health Research Center.
- Ishizaka, K., Marshall, S.P., & Conte, J.M. (2001). Individual differences in attentional strategies in multitasking situations. *Human Performance*, 14, 339-358.
- Joslyn, S., & Hunt, E. (1998) Evaluating individual differences in response to time-pressure situations. *Journal of Experimental Psychology: Applied*, 4, 16-43.
- Kahneman, D. (1973). *Attention and Effort*. Englewood Cliffs, J.J.: Prentice-Hall.
- Kahneman, D., Ben-Ishai, R., & Lotan, M. (1973). Relation of a test of attention to road accidents. *Journal of Applied Psychology*, 58, 113-115.
- Kane, M. (2002). The role of the prefrontal cortex in working memory capacity, executive attention, and general fluid intelligence. An individual differences perspective. *Psychonomic Bulletin and Review*, 9, 637-671.

- Kane, M., Bleckley, M.K., Conway, A., & Engle, R. (2001). A controlled-attention view of working-memory capacity. *Journal of Experimental Psychology: General*, 130, 169-183.
- Kane, M., & Engle, R. (2003). Working memory capacity and the control of attention: The contributions of goal neglect, response competition, and task set to Stroop interference. *Journal of Experimental Psychology: General*, 132, 47-70.
- Lansman, M., Poltrock, S., & Hunt, E. (1983). Individual differences in the ability to focus and divide attention. *Intelligence*, 7, 299-312.
- Larson, G.E., & Saccuzzo, D. (1989). Cognitive correlates of general intelligence: Toward a process theory of g. *Intelligence*, 13, 5-31.
- Law, D.J., Morrin, K.A., & Pellegrino, J.W. (1995). Training effects and working memory contributions to skill acquisition in a complex coordination task. *Learning and Individual Differences*, 7, 207-234.
- Lieberman, M.D., & Rosenthal, R. (2001). Why introverts can't always tell who likes them: Multitasking and nonverbal decoding. *Journal of Personality & Social Psychology*, 80, 294-310.
- Matthews, K., & Brunson, B. (1979). Allocation of attention and the Type A coronary-prone behavior pattern. *Journal of Personality & Social Psychology*, 37, 2081-2090.
- Mayr, U., Kliegl, R., & Krampe, R. T. (1996). Sequential and coordinative processing dynamics in figural transformations across the life span. *Cognition*, 59, 61-90.
- Meiran, N., & Marciano, H. (2002). Limitations in advance task preparation: Switching the relevant stimulus dimension in speeded same-different comparisons. *Memory & Cognition*, 30, 540-550.
- Meyer, D.E., & Kieras, D.E. (1997). A computational theory of executive cognitive processes and multiple-task performance: Part 1. Basic mechanisms. *Psychological Review*, 104, 3-65.
- Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex "frontal lobe" tasks: A latent variable analysis. *Cognitive Psychology*, 41, 49-100.
- Morrin, K., Law, D., & Pellegrino, J. (1994). Structural modeling of information coordination abilities: An evaluation and extension of the Yee, Hunt, and Pellegrino model. *Intelligence*, 19, 117-144.
- Morrin, K.A. (1996). Information processing characteristics in the performance of a real-time information coordination task: Working memory and attentional constraints. Unpublished doctoral dissertation, Vanderbilt University, Nashville, TN.
- Navon, D., & Gopher, D. (1979). On the economy of the human processing system. *Psychological Review*, 86, 214-255.
- Norman, D. A., & Bobrow, D. G. (1975). On data-limited and resource-limited processes. *Cognitive Psychology*, 7, 44-64. p81
- Norman, D. A., & Shallice, T. (1980). Attention to action: willed and automatic control of behavior. In R.J. Davidson, G. E. Schwartz, & D. Shapiro (eds.). *Consciousness and Self Regulation*. New York: Plenum.
- Paisley, T., & Mangan, G. (1988). Personality and conditioning with appetitive and aversive stimuli. *Personality and Individual Differences*, 9, 68-78.

- Pashler, H. (1989). Dissociations and dependences between speed and accuracy: Evidence for a two component theory of divided attention in simple tasks. *Cognitive Psychology*, 21, 469-514.
- Pashler, H. (1994). Dual task interference in simple tasks: Data and theory. *Psychological Bulletin*, 116, 220-244.
- Pashler, H. (2000). Task switching and multitask performance. In S. Monsell & J. Driver (Eds). *Control of cognitive processes: Attention and performance XVIII*. (pp. 275-423). Cambridge, MA, US: The MIT Press.
- Pashler, H., Johnston, J. C., & Ruthruff, E. (2001). Attention and performance. *Annual Review of Psychology*, 52, 629-651.
- Prinzel, L.J. DeVries, H. Freeman, F. G., & Mikulka, P. (2001) Examination of automation-induced complacency and individual difference variates. NASA Technical Memorandum #2001-211413.
- Prinzel, L.J., Pope, A. T. & Freeman, F. G. (2001). Application of physiological self-regulation and adaptive task allocation techniques for controlling operator hazardous states of awareness. NASA Technical Memorandum #2001-211015.
- Proctor, R., Wang, D., Pick, D. (1998). An empirical evaluation of the SYNWORK1 multiple-task work environment. *Behavior Research Methods, Instruments & Computers*, 30, 287-305.
- Rypma, B., Berger, J., D'Esposito, M. (2002). The influence of working-memory demand and subject performance on prefrontal cortical activity. *Journal of Cognitive Neuroscience*, 14, 721-731.
- Rovira, E., & Zinni, M. (2002) Information & Decision Uncertainty: Effects of Unreliable Automaton on Multi-task Performance & Workload. Paper presented at APA Midyear Symposium.
- Rubinstein, J., Meyer, D. E., & Evans, J. E. (2001). Executive control of cognitive processes in task switching. *Journal of Experimental Psychology: Human Perception and Performance*, 27, 763-797.
- Schumacher, E.H., Seymour, T.L., Glass, J.M., Fencsik, D.E., Lauber, E.J., Kieras, D.E., & Meyer, D.E. (2001). Virtually perfect time sharing in dual-task performance: uncorking the central cognitive bottleneck. *Psychological Science*, 12(2), 101-108.
- Shallice, T. & Burgess, P.W. (1991). Deficits in strategy application following frontal lobe damage in man. *Brain*, 114, 727-741.
- Sohn, M., & Anderson, J. (2001). Task preparation and task repetition: Two-component model of task switching. *Journal of Experimental Psychology: General*, 130, 764-778.
- Stankov, L. (1983). Attention and intelligence. *Journal of Educational Psychology*, 75, 471-490.
- Stuster, J. (2001). Salient human factors issues concerning the landing craft air cushion (LCAC). Anacapa Sciences, Inc.
- Stuster, J. W. (1991). Most salient human factors issues associated with Landing Craft Air Cushion (LCAC) crew positions. In *Proceedings of the LCAC Crew Medical Standards Conference*. Naval Health Research Center, San Diego, CA.

- Suss, H., Oberauer, K., Wittmann, W., Wilhelm, O., & Schulze, R. (2002). Working memory capacity explains reasoning ability —and a little bit more. *Intelligence*, 30, 261-288.
- Treisman, A. M. (1969). Strategies and models of selective attention. *Psychological Review*, 76, 282-292.
- Wickens, C. D. (1980). The structure of processing resources. In R. Nickerson & R. Pew (Eds.), *Attention and performance VIII*. Hillsdale NJ: Erlbaum.
- Wickens, C. (1984). Processing resources in attention. In R. Parasuraman & R. Davies (Eds.) *Varieties of attention*, (pp. 63-98). New York: Academic.
- Wickens, C.D. (1999). Cognitive factors in aviation. In F.T. Durso (Ed.) *Handbook of Applied Cognition* (pp. 247- 282). Chichester, England: John Wiley & Sons.
- Wilson, B, Evans, J., Emslie, H., Alderman, N., & Burgess, P. (1998). The development of an ecologically valid test for assessing patients with a dysexecutive syndrome. *Neuropsychological Rehabilitation*, 8, 213-228.
- Yee, P.L., Hunt, E., & Pellegrino, J.W. (1991). Coordinating cognitive information: Task effects and individual differences in integrating information from several sources. *Cognitive Psychology*, 23, 615-680.

APPENDIX A: SOURCES REVIEWED

- Ackerman, P. (1992). Predicting individual differences in complex skill acquisition: Dynamics of ability determinants. *Journal of Applied Psychology*, 77, 598-614.
- Ackerman, P., & Cianciolo, A. (2000). Cognitive, perceptual-speed, and psychomotor determinants of individual differences during skill acquisition. *Journal of Experimental Psychology: Applied*, 6, 259-290.
- Ackerman, P., & Cianciolo, A. (2002). Ability and task constraint determinants of complex task performance. *Journal of Experimental Psychology: Applied*, 8, 194-208.
- Ackerman, P., & Kanfer, R. (1993). Integrating laboratory and field study for improving selection: Development of a battery for predicting air traffic controller success. *Journal of Applied Psychology*, 78, 413-432.
- Ackerman, P., Beier, M., & Boyle, M. (2002). Individual differences in working memory within a nomological network of cognitive and perceptual speed abilities. *Journal of Experimental Psychology: General*, 131, 567-589.
- Ackerman, P., Kanfer, R., & Goff, M. (1995). Cognitive and noncognitive determinants and consequences of complex skill acquisition. *Journal of Experimental Psychology: Applied*, 1, 270-304.
- Adams, R. (2002). *Decision-making styles associated with accidents: Defining the high-risk pilot*. Unpublished Doctoral Dissertation.
- Anderson, J., Bothell, D., Douglass, S., Haimson, C., & Sohn, M-H. (2002). *CMU-ASP: ACT-R Learns to be an Anti-air warfare coordinator*. Department of the Navy, Office of Naval Research, <http://act-r.psy.cmu.edu/workshops/workshop-2002/talks/>.
- Baddeley, A. (1996). Exploring the central executive. *Quarterly Journal of Experimental Psychology*, 49A, 5-28.
- Baddeley, A. (2002). Is working memory still working? *European Psychologist*, 7, 85-97.
- Baddeley, A., Chincotta, D., & Adlam, A. (2001). Working memory and the control of action: Evidence from task switching. *Journal of Experimental Psychology: General*, 130, 641-657.
- Biggerstaff, S., Blower, D.J., Portman, C.A., & Chapman, A. (1998). Landing craft air cushion (LCAC) navigator selection system: Initial model development. *Naval Aerospace Medical Research Lab*, Pensacola Florida; NAMRL-1399, AD-A350-546.
- Bishop, K. (2001). *Working smart and working hard: The effects of entrepreneurial multi-tasking and intuitive activities on venture performance*. Unpublished Doctoral Dissertation.
- Brookings, J.B. (1990). A confirmatory factor analytic study of time sharing performance and cognitive abilities. *Intelligence*, 14, 43-59.
- Brunken, R., Steinbacher, S., Plass, J.L., & Leutner, D. (2002). Assessment of cognitive load in multimedia learning using dual-task methodology. *Experimental Psychology*, 49(2), 109-119.
- Burgess, P. (2000). Real world multi-tasking from a cognitive neuroscience perspective. In S. Monsell & J. Driver (Eds.), *Control of Cognitive Processes: Attention & Performance* (p. 465-472). Cambridge, MA, US: The MIT Press.

- Burgess, P. (2000). Strategy application disorder: The role of the frontal lobes in human multi-tasking. *Psychological Research*, 63, 279-288.
- Burgess, P.; Veitch, E., de Lacy Costello, A., Shallice, T. (2000). The cognitive and neuroanatomical correlates of multi-tasking. *Neuropsychologia*, 38, 848-863.
- Caldwell, J., Ramspott, S. (1998). Effects of task duration on sensitivity to sleep deprivation using the multi-attribute task battery. *Behavior Research Methods, Instruments & Computers*, 30, 651-660.
- Carlson, R.A.; Sohn, M-H.; Burgess, P.W. (2000). Control of multistep tasks. In Monsell, S., Driver, J., (Eds). *Control of cognitive processes: Attention and performance XVIII*. (pp. 425-472). Cambridge, MA, US: The MIT Press.
- Carlson, R.A.;& Sohn, M-H.; (2000). Cognitive control of multistep routines: Information processing and conscious intentions. In Monsell, S., Driver, J., (Eds). *Control of cognitive processes: Attention and performance XVIII*. (pp. 443-464). Cambridge, MA, US: The MIT Press.
- Carpenter, P.A., Just, M.A., & Reichle, E.D. (2000). Working memory and executive function: evidence from neuroimaging. *Current Opinion in Neurobiology* 10,195-199.
- Carpenter, P.A., Just, M.A., & Shell, P. (1990). What one intelligence test measures A theoretical account of the processing in the Raven progressive matrices test. *Psychological Review*, 97(3), 404-431.
- Conway, A., Cowan, N., Bunting, M., Theriault, D., & Minkoff, S. (2002). A latent variable analysis of working memory capacity, short-term memory capacity, processing speed, and general fluid intelligence. *Intelligence*, 30, 163-184.
- Craig, P. (1999). *Improving pilot decision-making in situations of high stakes, high stress, and time pressures*. Unpublished Doctoral Dissertation.
- Delbridge, K. (2001). *Individual differences in multi-tasking ability: Exploring a nomological network*. Unpublished Doctoral Dissertation.
- Dickman, S. (1990). Functional and dysfunctional impulsivity: Personality and cognitive correlates. *Personality and Individual Differences*, 58, 95-102.
- Dickman, S., & Meyer, D. (1988). Impulsivity and speed-accuracy tradeoffs in information processing. *Journal of Personality and Social Psychology*, 54, 274-290.
- Duncan, J., Burgess, P., & Emslie, H. (1995). Fluid intelligence after frontal lobe lesions. *Neuropsychologia*, 33, 261-268.
- Elsmore, T. (1994). SYNWORK1: A PC-based tool for assessment of performance in a simulated work environment. *Behavior Research Methods, Instruments & Computers*, 26, 421-426.
- Embretson, S. (1995). The role of working memory capacity and general control processes in intelligence. *Intelligence*, 20, 169-189.
- Emerson, M., & Miyake, A. (2003). The role of inner speech in task switching: A dual-task investigation. *Journal of Memory and Language*, 48, 148-168.
- Emerson, M., & Miyake, A., & Rettinger, D. (1999). Individual differences in integrating and coordinating multiple sources of information. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 25, 1300-1321.

- Engle, R., Tuholski, W., Laughlin, J., & Conway, A. (1999). Working memory, short-term memory, and general fluid intelligence: A latent variable approach. *Journal of Experimental Psychology: General*, 128, 309-331.
- Franklin, A.L., & Hunt, E. (1993). An emergency situation simulator for examining time-pressured decision making. *Behavior Research Methods, Instruments, & Computers*, 25, 143-147.
- Freed, M. (1998). Managing multiple tasks in complex, dynamic environments. In *Proceedings of the 1998 National Conference on Artificial Intelligence*, Madison, Wisconsin.
- Gopher, D., Armony, L., & Greenshpan, Y. (2000). Switching tasks and attention policies. *Journal of Experimental Psychology: General*, 129, 308-339.
- Gopher, D., Weil, M. & Bareket, T. (1994). Transfer of skill from a computer game trainer to flight. *Human Factors*, 36, 1-19.
- Griffin, G.R. & McBride, D.K. (1997). Multitask performance: Predicting success in naval aviation primary flight training. *Naval Aerospace Medical Research Lab*, Pensacola, Florida; AD-A168-246.
- Griffin, G.R. (1998). Predicting naval aviator flight training performance using multiple regression and an artificial neural network. *The International Journal of Aviation Psychology*, 8(2), 121-135.
- Gronlund, S., Ohrt, D., Dougherty, M., Perry, J., & Manning, C. (1998). Role of memory in air traffic control. *Journal of Experimental Psychology: Applied*, 4, 263-280.
- Hegarty, M., Shaw, P., & Miyake A. (2000). Constraints on using the dual-task methodology to specify the degree of central executive involvement in cognitive tasks. *Memory & Cognition*, 28 (3), 376-385.
- Humphreys, Glyn W.; Forde, Emer M. E.; & Francis, D. (2000). The organization of sequential actions. In Monsell, S., Driver, J., (Eds). *Control of cognitive processes: Attention and performance XVIII*. (pp. 427-442). Cambridge, MA, US: The MIT Press.
- Hunt, E., & Joslyn, S. (2000). A functional task analysis of time-pressured decision making. In J. Schraagen, S. Chipman, et al., (Eds) *Cognitive task analysis*. (pp. 119-132). Mahwah, NJ, US: Lawrence Erlbaum Associates, Publishers.
- Ishizaka, Kyoko; Marshall, Sandra P.; Conte, Jeffrey M. (2001). Individual differences in attentional strategies in multi-tasking situations. *Human Performance*, 14, 339-358.
- Johnson, N.E., Saccuzzo, D.P., & Guertin, T.L. (1994). The development and validation of a reliable alternate form for Raven's standard progressive matrices. *Assessment*, 1(3), 315-319.
- Joslyn, S., & Hunt, E. (1998). Evaluating individual differences in response to time-pressure situations. *Journal of Experimental Psychology: Applied*, 4, 16-43.
- Kane, M., & Engle, R. (2001). WM capacity, proactive interference, and divided attention: Limits on long-term memory retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 26, 336-358.
- Kane, M., & Engle, R. (2003). Working-memory capacity and the control of attention: The contributions of goal neglect, response competition, and task set to Stroop interference. *Journal of Experimental Psychology: General*, 132, 47-700.

- Kane, M., Bleckley, M., Conway, A., & Engle, R. (2001). A controlled-attention view of working-memory capacity. *Journal of Experimental Psychology: General*, 130, 169-183.
- Kieras, D.E. & Meyer, D.E. (1998). The EPIC architecture: Principles of operation. *EPIC Architecture* (5/1/98): <ftp.eecs.umich.edu/people/kieras/EPICArch.ps>.
- Kramer, Arthur F.; Hahn, Sowon; Gopher, Daniel. (1999). Task coordination and aging: Explorations of executive control processes in the task switching paradigm. *Acta Psychologica*, 101, 339-378.
- Kyllonen, P. (1996). Is working memory capacity Spearman's g?. In Dennis, I. & Tapsfield, P., (Eds). *Human Abilities, Their Nature and Measurement*, (pp. 49-75). Mahwah, NJ, Lawrence Erlbaum Associates.
- Kyllonen, P., & Christal, R. (1990). Reasoning ability is (little more than) working-memory capacity? *Intelligence*, 14, 389-433.
- Larson, G., & Saccuzzo, D. (1989). Cognitive correlates of general intelligence: Toward a process theory of g. *Intelligence*, 13, 5-31.
- Lieberman, Matthew D.; Rosenthal, Robert. (2001). Why introverts can't always tell who likes them: Multi-tasking and nonverbal decoding. *Journal of Personality & Social Psychology*, 80, 294-310.
- Lin, D., & Su, Y. (1998). The effect of time pressure on expert system based training for emergency management. *Behaviour & Information Technology*, 17, 195-202.
- Lohman, D.F. (1996). Spatial ability and g. In Dennis, I. & Tapsfield, P., (Eds). *Human Abilities, Their Nature and Measurement*, (pp. 97-116). Mahwah, NJ, Lawrence Erlbaum Associates.
- Maule, A.J., Hockey, G.R.J., Bdzola, L. (2000). Effects of time-pressure on decision-making under uncertainty: changes in affective state and information processing strategy. *Acta Psychologica*, 104, 283-301.
- Mecklinger, A., Von Cramon, D., Springer, A., & Matthes-Von Cramon, G. (1999). Executive control functions in task switching: Evidence from brain injured patients. *Journal of Clinical and Experimental Neuropsychology*, 21, 606-619.
- Meiran, N., & Marciano, H. (2002). Limitations in advance task preparation: Switching the relevant stimulus dimension in speeded same-different comparisons. *Memory & Cognition*, 30, 540-550.
- Meyer, D. E. & Kieras, E. D. (1997a). A computational theory of executive cognitive processes and multiple task performance: Part 1. Basic mechanisms. *Psychological Review*, 104-4, 3-75.
- Meyer, D., Glass, J., Mueller, S., Seymour, T., & Kieras, D. (2001) Executive-process interactive control: A unified computational theory for answering 20 questions (and more) about cognitive ageing. *European Journal of Cognitive Psychology*, 13, 123-164.
- Meyer, D.E., & Kieras, D.E. (1997b). A computational theory of executive cognitive processes and multiple-task performance: Part 2. Accounts of Psychological Refractory-Period Phenomena. *Psychological Review*, 104-4, 749-791.
- Miyake, A., & Shah, P. (1999). *Models of working memory: Mechanisms of active maintenance and executive control*. New York, NY, US: Cambridge University Press.

- Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex "frontal lobe" tasks: A latent variable analysis. *Cognitive Psychology*, 41, 49-100.
- Miyake, A., Friedman, N., Rettinger, D., Shah, P., & Hegarty, M. (2001). How are visuospatial working memory, executive functioning, and spatial abilities related? A latent-variable analysis. *Journal of Experimental Psychology: General*, 130, 621-640.
- Nontasak, T. (1997). Landing craft air cushion (LCAC) vehicle operator/engineer selection research activities. *Naval Aerospace Medical Research Lab*, Pensacola, Florida; AD-A202-613.
- Pashler, H., (2000). Task switching and multitask performance. In S. Monsell & J. Driver (Eds). *Control of cognitive processes: Attention and performance XVIII*. (pp. 275-423). Cambridge, MA, US: The MIT Press.
- Pashler, H., Johnston, J. C., & Ruthruff, E. (2001). Attention and performance. *Annual Review of Psychology*, 52, 629-651.
- Pesante, J., Williges, R., Woldstad, J. (2001). The effects of multi-tasking on quality inspection in advanced manufacturing systems. *Human Factors & Ergonomics in Manufacturing*, 11, 287-298.
- Prinzel, L.J., DeVries, H., Freeman, F. G., & Mikulka, P. (2001) Examination of automation-induced complacency and individual difference variates. NASA Technical Memorandum #2001-211413.
- Prinzel, L.J., Pope, A. T. & Freeman, F. G. (2001). Application of physiological self-regulation and adaptive task allocation techniques for controlling operator hazardous states of awareness. NASA Technical Memorandum #2001-211015.
- Proctor, R., Wang, D., Pick, D. (1998). An empirical evaluation of the SYNWORK1 multiple-task work environment. *Behavior Research Methods, Instruments & Computers*, 30, 287-305.
- Rubinstein, J., Meyer, D. E., & Evans, J. E. (2001). Executive control of cognitive processes in task switching. *Journal of Experimental Psychology: Human Perception and Performance*, 27(4), 763-797.
- Salthouse, T., & Miles, J. (2002). Aging and time-sharing aspects of executive control. *Memory & Cognition*, 30, 572-582.
- Salthouse, T., Fristoe, N., McGuthry, K., & Hambrick, D. (1998). Relation of task switching to speed, age, and fluid intelligence. *Psychology & Aging*, 13, 445-461.
- Salthouse, T., Hambrick, D. Lukas, K., & Dell, T. (1996). Determinants of adult age differences on synthetic work performance. *Journal of Experimental Psychology: Applied*, 2, 305-329.
- Schumacher, E., Seymour, T., Glass, J., Fencsik, D., Lauber, E., Kieras, D., & Meyer, D. (2001). Virtually perfect time sharing in dual-task performance: Uncorking the central cognitive bottleneck. *Psychological Science*, 121, 101-108.
- Schunn, C., & Reder, L. (2001). Another source of individual differences: Strategy adaptivity to changing rates of success. *Journal of Experimental Psychology: General*, 130, 59-76.
- Scott, E. & Marshall, S.P. A hybrid model of novices' performance on a simulated CIC task. (4/23/03): http://act-r.psy.cmu.edu/workshops/Workshop-98/Scott/quick_index.html.

- Shah, P., & Miyake, A. (1996). The separability of working memory resources for spatial thinking and language processing: An individual differences approach. *Journal of Experimental Psychology: General*, 125, 4-27.
- Shallice, T. & Burgess, P.W. (1991). Deficits in strategy application following frontal lobe damage in man. *Brain*, 114, 727-741.
- Sohn, M., & Anderson, J. (2001). Task preparation and task repetition: Two-component model of task switching. *Journal of Experimental Psychology: General*, 130, 764-778.
- Stuster, J. W. (1991). Most salient human factors issues associated with Landing Craft Air Cushion (LCAC) crew positions. In, *Proceedings of the LCAC Crew Medical Standards Conference*. Naval Health Research Center, San Diego, CA.
- Stuster, J.W. (2001). Salient human factors issues concerning the landing craft air cushion (LCAC). Anacapa Sciences, Inc.
- Suss, H., Oberauer, K., Wittmann, W., Wilhelm, O., & Schulze, R. (2002). Working memory capacity explains reasoning ability—and a little bit more. *Intelligence*, 30, 261-288.
- Tsang, P. (1998). Age, attention, expertise, and time-sharing performance. *Psychology & Aging*, 13, 323-347.
- Verhaeghen, P. & Cerella, J. (2002). Aging, executive control, and attention: a review of meta-analyses. *Neuroscience and Biobehavioral Reviews*, 26, 849-857.
- Wilson, B., Evans, J., Emslie, H. Alderman, N., & Burgess, P. (1998). The development of an ecologically valid test for assessing patients with a dysexecutive syndrome. *Neuropsychological Rehabilitation*, 8, 213-228.
- Woltz, D.J. (1988). An investigation of the role of working memory in procedural skill acquisition. *Journal of Experimental Psychology: General*, 117, 319-331.
- Woltz, D.J., Gardner, M.K., Gyll, S.P. (2000). The role of attention processes in near transfer of cognitive skills. *Learning and Individual Differences*, 12, 209-251.
- Wylie, G. & Allport, A. (2000). Task switching and the measurement of "switch costs". *Psychological Research*, 63, 212-233.
- Xie, B. & Salvendy, G. (2000). Prediction of mental workload in single and multiple tasks environments. *International Journal of Cognitive Ergonomics*, 4(3), 213-242.
- Xie, B. (1997). A methodology for modeling and predicting mental workload in single- and multi-task environments. Unpublished Doctoral Dissertation.
- Xie, B., & Salvendy, G. (2000). Review and reappraisal of modeling and predicting mental workload in single- and multi-task environments. *Work & Stress*, 14, 74-99.
- Yee, P.L., Hunt, E., & Pellegrino, J.W. (1991). Coordinating cognitive information: Task effects and individual differences in integrating information from several sources. *Cognitive Psychology*, 23, 615-680.

APPENDIX B. SUMMARY OF STANDARDS

VALIDITY STANDARDS

- 1.1 A rationale should be presented for each recommended interpretation and use of test scores, together with a comprehensive summary of the evidence and theory bearing on the intended use or interpretation.
- 1.2 The test developer should set forth clearly how test scores are intended to be interpreted and used. The population for which a test is appropriate should be clearly delimited and the construct that the test is intended to assess should be clearly described,
- 1.3 If validity for some common or likely interpretation has not been investigated, or if the interpretation is inconsistent with available evidence, that fact should be made clear and potential users should be cautioned about making unsupported claims
- 1.4 If a test is used in a way that has not been validated, it is incumbent on the user to justify the new use, collecting new evidence if necessary.
- 1.5 The composition of any sample of examinees from which validity evidence is obtained should be described in as much detail as is practical, including major relevant sociodemographic and developmental characteristics
- 1.6 When the validation rests in part on the appropriateness of test content, the procedures followed in specifying and generating test content should be described and justified in reference to the construct the test is intended to measure or the domain it is intended to represent. If the definition of the content sample incorporates criteria such as importance, frequency, or criticality, these should also be clearly explained and justified.
- 1.7 When a validation rests in part on the opinions or decisions of expert judges, observers, or raters, procedures for selecting such experts and for eliciting judgments or ratings should be fully described.
- 1.8 If the rationale for a test use or score interpretation depends on premises about the psychological processes or cognitive operations used by examinees, then theoretical or empirical evidence in support of those premises should be provided. When statements about the processes employed by observers or scorers are part of the argument for validity, similar information should be provided.
- 1.9 If a test is claimed to be essentially unaffected by practice and coaching, then the sensitivity of test performance to change with these forms of instruction should be documented.
- 1.10 When interpretation of performance on specific items, or small subsets of items is suggested, the rationale and relevant evidence in support of such interpretation should be provided. When interpretation of individual item responses is likely but

is not recommended by the developer, the user should be warned against making such interpretations.

- 1.11 If the rationale for a test use or interpretation on premises about the relationships among parts of the test, evidence concerning the internal structure of the test should be provided.
- 1.12 When interpretation of subscores, score differences, or profiles is suggested, the rationale and relevant evidence in support of such interpretation should be provided. Where composite scores are developed, the basis and rationale for arriving at the composites should be given.
- 1.13 When validity evidence includes statistical analyses of test results, either alone or together with data on other variables, the conditions under which the data were collected should be described in enough detail that users can judge the relevance of the statistical findings to local conditions. Attention should be drawn to any features of a validation data collection that are likely to differ from typical operational testing conditions and that could plausibly influence test results.
- 1.14 When validity evidence includes empirical analyses of test responses together with data on other variables, the rationale for selecting the additional variables should be provided. Where appropriate and feasible, evidence concerning the constructs represented by other variables, as well as their technical properties, should be presented or cited. Attention should be drawn to any likely sources of dependence (or lack of dependence) among variables other than dependences among the construct they represent.
- 1.15 When it is asserted that a certain level of test performance predicts adequate or inadequate criterion performance, information about the levels of criterion performance associated with given levels of test scores should be provided.
- 1.16 When validation relies on evidence that test scores are related to one or more criterion variables, information about the suitability and technical quality of the criteria should be reported
- 1.17 If test scores are used in conjunction with other quantifiable variables to predict some outcome or criterion, regression (or equivalent) analyses should include those additional relevant variables along with the test scores.
- 1.18 When statistical adjustments, such as those for restriction of range or attenuation, are made, both adjusted and unadjusted coefficients, as well as the specific procedure used, and all statistics used in the adjustment, should be reported
- 1.19 If a test is recommended for use in assigning persons to alternative treatments or is likely to be so used, and if outcomes from those treatments can reasonably be compared on a common criterion, then, whenever feasible, supporting evidence of differential outcomes should be provided.
- 1.20 When a meta-analysis is used as evidence of the strength of a test-criterion relationship, the test and criterion variables in the local situation should be comparable with those in the studies summarized. If relevant research includes credible evidence that any other features of the testing application may influence

the strength of the test-criterion relationship, the correspondence between those features in the local situation and in the meta-analysis should be reported. Any significant disparities that might limit the applicability of the meta-analytic findings to the location situation should be noted explicitly.

- 1.21 Any meta-analytic evidenced used to support an intended test use should be clearly described, including methodological choices in identifying and coding studies, correcting for artifacts, and examining potential moderator variables. Assumptions made in correcting for artifacts such as criterion unreliability and range restriction should be presented, and the consequences of these assumptions made clear.
- 1.22 When it is clearly stated or implied that a recommended test use will result in a specific outcome, the basis for expecting that outcome should be presented together with relevant evidence.
- 1.23 When a test use of score interpretation is recommended on the grounds that testing or the testing program per se will result in some indirect benefit in addition to the utility of information from the test scores themselves, the rationale for anticipating the indirect benefit should be made explicit. Logical or theoretical arguments and empirical evidence for the indirect benefit should be provided. Due weight should be given to any contradictory findings in the scientific literature, including findings suggesting important indirect outcomes other than those predicted.
- 1.24 When unintended consequences result from test use, an attempt should be made to investigate whether such consequences arise from the test's sensitivity to characteristics other than those it is intended to assess or to the test's failure fully to represent the intended construct.

RELIABILITY AND ERRORS OF MEASUREMENT STANDARDS

- 2.1 For each total score, subscore, or combination of scores that is to be interpreted, estimates of relevant reliabilities and standard errors of measurement or test information functions should be reported.
- 2.2 The standard error of measurement, both overall and conditional (if relevant) should be reported both in raw score or original scale units and in units of each derived score recommended for use in test interpretation.
- 2.3 When test interpretation emphasizes differences between two observed scores of an individual or two averages of a group, reliability data, including standard errors, should be provided for such differences.
- 2.4 Each method of quantifying the precision or consistency of scores should be described clearly and expressed in terms of statistics appropriate to the method. The sampling procedures used to select examinees for reliability analyses and descriptive statistics on these samples should be reported.
- 2.5 A reliability coefficient or standard error of measurement based on one approach should not be interpreted as interchangeable with another derived by a different technique unless their implicit definitions of measurement error are equivalent.

- 2.6 If reliability coefficients are adjusted for restriction of range or variability, the adjustment procedure and both the adjusted and unadjusted coefficients should be reported. The standard deviations of the group actually tested and of the target population, as well as the rationale for the adjustment, should be presented.
- 2.7 When subsets of items within a test are dictated by the test specifications and can be presumed to measure partially independent traits or abilities, reliability estimation procedures should recognize the multifactor character of the instrument.
- 2.8 Test users should be informed about the degree to which rate of work may affect examinee performance.
- 2.9 When a test is designed to reflect rate of work, reliability should be estimated by the alternate-form or test-retest approach, using separately timed administrations.
- 2.10 When subjective judgment enters into test scoring, evidence should be provided on both inter-rater consistency in scoring and within-examinee consistency over repeated measurements. A clear distinction should be made among reliability data based on (a) independent panels of raters scoring the same performances or products, (b) a single panel scoring successive performances or new products, and (c) independence panels scoring successive performances or new products.
- 2.11 If there are generally accepted theoretical or empirical reasons for expecting that reliability coefficients, standard errors of measurement, or test information functions will differ substantially for various subpopulations, publishers should provide reliability data as soon as feasible for each major population for which the test is recommended.
- 2.12 If a test is proposed for use in several grades or over a range of chronological age groups and if separate norms are provided for each grade or each age group, reliability data should be provided for each age or grade population, not solely for all grades or ages combined.
- 2.13 If local scorers are employed to apply general score rules and principles specified by the test developer, local reliability data should be gathered and reported by local authorities when adequate size samples are available.
- 2.14 Conditional standard errors of measurement should be reported at several score levels if constancy cannot be assumed. Where cut scores are specified for selection or classification, the standard errors of measurement should be reported in the vicinity of each cut score.
- 2.15 When a test or combination of measures is used to make categorical decisions, estimates should be provided of the percentage of examinees who would be classified in the same way on two applications of the procedure, using the same form or alternate forms of the instrument.
- 2.16 In some testing situations, the items vary from examinee to examinee—through random selection from an extensive item pool or application of algorithms based on the examinee's level of performance on previous items or preferences with respect to item difficulty. In this type of testing the preferred approach to reliability

estimation is one based on successive administrations of the test under conditions similar to those prevailing in operational test use.

- 2.17 When a test is available in both long and short versions, reliability data should be reported for scores on each version, preferably based on an independent administration of each.
- 2.18 When significant variations are permitted in test administration procedures, separate reliability analyses should be provided for scores produced under each major variation if adequate sample sizes are available.
- 2.19 When average test scores for groups are used in program evaluations, the groups tested should generally be regarded as a sample from a larger population, even if all examinees available at the time of measurement are tested. In such cases the standard error of the group mean should be reported, as it reflects variability due to sampling of examinees as well as variability due to measurement error.
- 2.20 When the purpose of testing is to measure the performance of groups rather than individuals, a procedure frequently used is to assign a small subset of items to each of many subsamples of examinees. Data are aggregated across subsamples and item subsets to obtain a measure of group performance. When such procedures are used for program evaluation or population descriptions, reliability analyses must take the sampling scheme into account.

TEST DEVELOPMENT AND REVISION STANDARDS

- 3.1 Tests and testing programs should be developed on a sound scientific basis. Test developers and publishers should compile and document adequate evidence bearing on test development.
- 3.2 The purpose(s) of the test, definition of the domain, and the test specifications should be stated clearly so that judgments can be made about the appropriateness of the defined domain for the stated purpose(s) of the test and about the relation of items to the dimensions of the domain they are intended to represent.
- 3.3 The test specifications should be documented, along with their rationale and the process by which they were developed. The test specifications should define the content of the test, the proposed number of items, the item formats, the desired psychometric properties of the items, and the item and section arrangement. They should also specify the amount of time of testing, directions to the test takers, procedures to be used for test administration and scoring, and other relevant information.
- 3.4 The procedures used to interpret test scores, when appropriate, the normative or standardization samples, or the criterion used should be documented.
- 3.5 When appropriate, relevant experts external to the testing program should review the test specifications. The purpose of the review, the process by which the review is conducted, and the results of the review should be documented. The qualifications, relevant experience and demographic characteristics of expert judges should be documented.

- 3.6 The type of items, the response formats, scoring procedures, and test administration procedures should be selected based on the purposes of the test, the domain to be measured, and the intended test takers.
- 3.7 The procedures used to develop, review, and try out items, and to select items from the item pool should be documented. If the items were classified into different categories of subtests according to the test specifications the procedures used for the classification and the appropriateness and accuracy of the classification should be documented.
- 3.8 When item tryouts or field tests are conducted, the procedures used to select the sample of test takers for item tryouts and the resulting characteristics of the sample should be documented. When appropriate the sample should be as representative as possible of the population for which the test is intended
- 3.9 When a test developer evaluates the psychometric properties of items, the classical or item response theory model used for evaluating the psychometric properties of items should be documented.
- 3.10 Test developers should conduct cross validation studies when items are selected primarily on the basis of empirical relationships rather than on the basis of content or theoretical considerations
- 3.11 Test developers should document the extent to which the content domain of a test represents the defined domain and test specifications
- 3.12 The rationale and supporting evidence for computerized adaptive tests should be documented.
- 3.13 When a test score is derived from the differential weighting of items, the test developer should document the rationale and process used to develop, review, and assign item weights
- 3.14 The criteria used for scoring test takers' performance on extended response items should be documented. This documentation is especially important for performance assessments such as scorable portfolios and essays where the criteria for scoring may not be obvious to the user.
- 3.15 When using a standardized testing format to collect structured behavior samples, the domain, test design, test specifications and materials should be documented as for any other test.
- 3.16 If a short form of a test is prepared for example by reducing the number of items on the original test or organizing portions of a test into a separate form, the specs of the short form should be as similar as possible to those of the original test.
- 3.17 When previous research indicates that irrelevant variance should confound the domain definition underlying the test, then to the extent feasible, the test developer should investigate sources of irrelevant variance.
- 3.18 For tests that have time limits, test development research should examine the degree to which scores include a speed component and evaluate the

appropriateness of that component, given the domain the test is designed to measure.

- 3.19 The directions for test administration should be presented with sufficient clarity and emphasis so that it is possible for others to replicate adequately the administration conditions under which the data on reliability and validity and where appropriate norms were obtained.
- 3.20 The instructions presented to test takers should contain sufficient detail so that test takers can respond to a task in the manner that the test developer intended. When appropriate, sample material, practice or sample questions, criteria for scoring, and a representative item identified with each major area in the test's classification or domain should be provided to the test takers prior to the administration of the test or included in the testing material as part of the standard administration instructions.
- 3.21 If the test developer indicates that the conditions of administration are permitted to vary from one test taker or group to another permissible variation in conditions for administration should be identified
- 3.22 Procedures for scoring and, if relevant, scoring criteria should be presented by the test developer in sufficient detail and clarity to maximize the accuracy of scoring. Instructions for using rating scales or for deriving scores obtained by coding, scaling, or classifying constructed responses should be clear.
- 3.23 The process for selecting, training, and qualifying scorers should be documented by the test developer.
- 3.24 When scoring is done locally and requires scorer judgment, the test user is responsible for providing adequate training and instruction to the scorers and for examining score agreement and accuracy.
- 3.25 A test should be amended or revised when new research data, significant changes in the domain represented, or newly recommended conditions of test use may lower the validity of test score interpretations.
- 3.26 Test should be labeled or advertised and revised only when they have been revised in significant ways.
- 3.27 If a test or part of a test is intended for research only and is not distributed for operational use, statements to this effect should be displayed prominently on all relevant test administration and interpretation materials that are provided to the test user. [all tests given to participants in the validity studies will be labeled for research only]

SCALES, NORMS, AND SCORE COMPARABILITY STANDARDS

- 4.1 Test documents should provide test users with clear explanations of the meaning and intended interpretation of derived score scales, as well as their limitations.
- 4.2 The construction of scales used for reporting scores should be described clearly in test documents.

- 4.3 If there is a sound reason to believe that specific misinterpretations of a score scale are likely, test users should be explicitly forewarned.
- 4.4 When raw scores are intended to be directly interpretable, their meanings, intended interpretations, and limitations should be described and justified in the same manner as is done for derived score scales.
- 4.5 Norms, if used, should refer to clearly described populations. These populations should include individuals or groups to whom test users will ordinarily wish to compare their own examinees.
- 4.6 Reports of norming studies should include precise specification of the population that was sampled, sampling procedures and participation rates, and descriptive statistics. The information provided should be sufficient to enable users to judge the appropriateness of the norms for interpreting the scores of local examinees. Technical documentation should indicate the precision of the norms themselves.
- 4.7 If local examinee groups differ materially from the populations to which norms refer, a user who reports derived scores based on the published norms has the responsibility to describe such differences if they bear upon the interpretation of the reported scores.
- 4.8 When norms are used to characterize examinee groups, the statistics used to summarize each group's performance and the norms to which those statistics are referred should be clearly defined and should support the intended use or interpretation.
- 4.9 When raw score or derived score scales are designed for criterion-referenced interpretation, including the classification of examinees into separate categories, the rationale for recommended score interpretations should be clearly explained.
- 4.10A clear rationale and supporting evidence should be provided for any claim that scores earned on different forms of a test may be used interchangeably. In some cases, direct evidence of score equivalence may be provided. In other cases, evidence may come from a demonstration that the theoretical assumptions underlying procedures for establishing score comparability have been sufficiently satisfied. The specific rationale and the evidence required will depend in part on the intended uses for which score equivalence is claimed.
- 4.11When claims of form-to-form score equivalence are based on equating procedures, detailed technical information should be provided on the method by which equating functions or other linkages were established and on the accuracy of equating functions.
- 4.12In equating studies that rely on the statistical equivalence of examinee groups receiving different forms, methods of assuring such equivalence should be described in detail.
- 4.13In equating studies that employ an anchor test design, the characteristics of the anchor test and its similarity to the forms being equated should be presented, including both content specifications and empirically determined relationships among test scores. If anchor items are used, as in some IRT-based and classical

equating studies, the representativeness and psychometric characteristics of anchor items should be presented.

- 4.14 When score conversions or comparison procedures are used to relate scores on tests or test forms that are not closely parallel, the construction, intended interpretation, and limitations of those conversions or comparisons should be clearly described.
- 4.15 When additional test forms are created by taking a subset of items in an existing test form or by rearranging its items and there is sound reason to believe that scores on these forms may be influenced by item context effects, evidence should be provided that there is no undue distortion of norms for the different versions or of score linkages between them.
- 4.16 If test specifications are changed from one version of a test to a subsequent version, such changes should be identified in the test manual, and an indication should be given that converted scores for the two versions may not be strictly equivalent. When substantial changes in test specifications occur, either scores should be reported on a new scale or a clear statement should be provided to alert users that the scores are not directly comparable with those on earlier versions of the test.
- 4.17 Testing programs that attempt to maintain a common scale over time should conduct periodic checks of the stability of the scale on which scores are reported.
- 4.18 If a publisher provides norms for use in test score interpretation, then so long as the test remains in print, it is the publisher's responsibility to assure that the test is renormed with sufficient frequency to permit continued accurate and appropriate test interpretations.
- 4.19 When proposed score interpretations involve one or more cut scores, the rationale and procedures used for establishing cut scores should be clearly documents.
- 4.20 When feasible, cut scores defining categories with distinct substantive interpretations should be established on the basis of sound empirical data concerning the relation of test performance to relevant data.
- 4.21 When cut scores defining pass-fail or proficiency categories are based on direct judgments about the adequacy of items or test performances or performance levels, the judgmental process should be designed so that judges can bring their knowledge and experience to bear in a reasonable way.

TEST ADMINISTRATION, SCORING, AND REPORTING STANDARDS

- 5.1 Test administrators should follow carefully the standardized procedures for administration and scoring specified by the test developer, unless the situation or a test taker's disability dictates that an exception should be made.
- 5.2 Modifications or disruptions of standardized test administration procedures or scoring should be documented.
- 5.3 When formal procedures have been established for requesting and receiving accommodations, test takers should be informed of these procedures in advance of testing.

- 5.4 The testing environment should furnish reasonable comfort with minimal distractions.
- 5.5 Instructions to test takers should clearly indicate how to make responses. Instructions should also be given in the use of any equipment likely to be unfamiliar to test takers. Opportunity to practice responding should be given when equipment is involved, unless use of the equipment is being assessed.
- 5.6 Reasonable efforts should be made to assure the integrity of test scores by eliminating opportunities for test takers to attain scores by fraudulent means.
- 5.7 Test users have the responsibility of protecting the security of test materials at all times.
- 5.8 Test scoring services should document the procedures that were followed to assure accuracy of scoring. The frequency of scoring errors should be monitored and reported to users of the service on reasonable request. Any systematic source of scoring errors should be corrected.
- 5.9 When test scoring involves human judgment, scoring rubrics should specify criteria for scoring. Adherence to established scoring criteria should be monitored and checked regularly. Monitoring procedures should be documented.
- 5.10 When test score information is released to students, parents, legal representatives, teachers, clients, or the media, those responsible for testing programs should provide appropriate interpretations. The interpretations should describe in simple language what the test covers, what scores mean, the precision of the scores, common misinterpretations of test scores, and how scores will be used.
- 5.11 When computer-prepared interpretations of test response protocols are reported, the sources, rationale, and empirical basis for these interpretations should be available, and their limitations should be described.
- 5.12 When group-level information is obtained by aggregating the results of partial tests taken by individuals, validity and reliability should be reported for the level of aggregation at which results are reported. Scores should not be reported for individuals unless the validity, comparability, and reliability of such scores have been established.
- 5.13 Transmission of individually identified test scores to authorized individuals should be done in a manner that protects the confidential nature of the scores.
- 5.14 When a material error is found in test scores or other important information released by a testing organization or other institution, a corrected score report should be distributed as soon as practicable to all known recipients who might otherwise use the erroneous scores as a basis for decision making. The corrected report should be labeled as such.
- 5.15 When test data about a person are retained, both the test protocol and any written report should also be preserved in some form. Test users should adhere to the policies and record-keeping practice of their professional organizations.

- 5.16 Organizations that maintain test scores on individuals in data files or in an individual's records should develop a clear set of policy guidelines on the duration of retention of an individual's records, and on the availability, and use over time, of such data.

SUPPORTING DOCUMENTATION FOR TESTS STANDARDS

- 6.1 Test documents (e.g., test manuals, technical manuals, user's guides, and supplemental material) should be made available to prospective test users and other qualified persons at the time a test is published or released for use.
- 6.2 Test documents should be complete, accurate, and clearly written so that the intended reader can readily understand the content.
- 6.3 The rationale for the test, recommended uses of the test, support for such uses, and information that assists in score interpretation should be documented. When particular misuses of a test can be reasonably anticipated, cautions against such misuses should be specified.
- 6.4 The population for whom the test is intended and the test specifications should be documented. If applicable, the item pool and scale development procedures should be described in the relevant test manuals. If normative data are provided, the norming population should be described in terms of relevant demographic variables, and the year(s) in which the data were collected should be reported.
- 6.5 When statistical descriptions and analyses that provide evidence of the reliability of scores and the validity of their recommended interpretations are available, the information should be included in the test's documentation. When relevant for test interpretation, test documents ordinarily should include item level information, cut scores and configural rules, information about raw scores and derived scores, normative data, the standard errors of measurement, and a description of the procedures used to equate multiple forms.
- 6.6 When a test relative to a course of training or study, a curriculum, a textbook, or packaged instruction, the documentation should include an identification and description of the course or instructional material and should indicate the year in which these materials were prepared.
- 6.7 Test documents should specify qualifications that are required to administer a test and to interpret the test scores accurately.
- 6.8 If a test is designed to be scored or interpreted by test takers, the publisher and test developer should provide evidence that the test can be accurately scored or interpreted by the test takers. Tests that are designed to be scored and interpreted by the test taker should be accompanied by interpretive materials that assist the individual in understanding the test scores and that are written in language that the test taker can understand.
- 6.9 Test documents should cite a representative set of the available studies pertaining to general and specific uses of the test.

- 6.10 Interpretive materials for tests, that include case studies, should provide examples illustrating the diversity of prospective test takers.
- 6.11 If a test is designed so that more than one method can be used for administration or for recording responses—such as marking responses in a test booklet, on a separate answer sheet, or on a computer keyboard—then the manual should clearly document the extent to which scores arising from these methods are interchangeable. If the results are not interchangeable, this fact should be reported, and guidance should be given for the interpretation of scores obtained under the various conditions or methods of administration.
- 6.12 Publishers and scoring services that offer computer-generated interpretations of test scores should provide a summary of the evidence supporting the interpretations given.
- 6.13 When substantial changes are made to a test, the test's documentation should be amended, supplemented, or revised to keep information for users current and to provide useful additional information or cautions.
- 6.14 Every test form and supporting document should carry a copyright date or publication date.
- 6.15 Test developers, publishers, and distributors should provide general information for test users and researchers who may be required to determine the appropriateness of an intended test use in a specific context. When a particular test user cannot be justified, the response to an inquiry from a prospective test user should indicate this fact clearly. General information also should be provided for test takers and legal guardians who must provide consent prior to a test's administration.