INITIAL VALIDATION OF NOVEL PERFORMANCE-BASED MEASURES: MENTAL ROTATION AND PSYCHOMOTOR ABILITY

by

PHILIP GEORGE FATOLITIS
M.A. University of West Florida, 2001
B.A. Chaminade University, 1996

A dissertation submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy
in the Department of Psychology
in the College of Sciences
at the University of Central Florida
Orlando, Florida

Spring Term
2008

Major Professors: Peter Hancock and Florian Jentsch

ABSTRACT

Given the high-risk nature of military flight operations and the significant resources required to train U.S. Naval Aviation personnel, continual improvement is required in the selection process. In addition to general commissioning requirements and aeromedical standards, the U.S. Navy utilizes the Aviation Selection Test Battery (ASTB) to select commissioned aviation students. Although the ASTB has been a good predictor of aviation student performance in training, it was proposed that incremental improvement could be gained with the introduction of novel, computer administered performance-based measures: Block Rotation (BRT) and a Navy-developed Compensatory Tracking task. This work constituted an initial validation of the BRT, an interactive virtual analog of Shepard-Metzler's (1971) Mental Rotation task that was developed with the intention of quantifying mental rotation and psychomotor ability. For Compensatory Tracking, this work sought to determine if data gathered concord with results in extant literature, confirming the validity of the task. Data from the BRT were examined to determine task reliability and to formulate relevant quantitative/predictive performance human models. Results showed that the BRT performance is a valid spatial ability predictor whose output can be modeled, and that Compensatory Tracking task data concord with the psychometric properties of tracking tasks that have been previously presented in the literature.

This dissertation is dedicated to my loving, patient, helpful and otherwise diligent wife, Lyn, and to my children, Georgia, Helena, Damon and those yet to come, who make all work worth doing.

ACKNOWLEDGMENTS

First, the United States Navy provided the opportunities to complete all of my post-secondary education. I am indebted to the Navy for this, but more importantly for providing access to the men and women of the Medical Corps, Nurse Corps, and Medical Service Corps who have been such an inspiration, in particular, those who are and have been associated with the aeromedical community. I am grateful to have been a student of Captain Sean Biggerstaff, without whose support and guidance I would not have been granted the opportunity to earn a Ph.D via the Navy. I am also grateful to Dr. Anil Raj of the Institute for Human and Machine Cognition for introducing the notion that "machinery" can be useful in the real-time diagnosis and enhancement of human performance. Captain John Schmidt and many others from the Naval Aerospace and Operational Medicine Institutes were a motivating force in the conduct of this research and I am grateful to them for having been introduced to this topic.

As for academic mentors, Dr. Peter Hancock is beyond compare. He is one of the finest instructors that any student could ask for: he brought humor and amicability to the explication of philosophy. In addition to providing the highest quality of academic instruction, Dr. Florian Jentsch helped me to navigate aspects of the extensive university maze and facilitated my ability to actually get some work done. Dr. Jentsch was more than generous with his time and talents and I hope that this work will do justice to the effort that he has invested. To have worked with Dr. Robert Kennedy was truly an honor. Having one of the most accomplished professionals in our field spend so much time with me made learning a privilege. Dr. Clint Bowers was also a great help in the execution of this work and I am grateful that he chose to be a part of this team.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1: INTRODUCTION

Problem Statement

Notwithstanding the high-risk nature of U.S. Naval Aviation training and operations, the Navy spends approximately $1 million to train each Naval Aviator and Naval Flight Officer from the time of accession to the time of winging, i.e., to the completion of Advanced Flight Training (about 3 years). This substantial cost underlies the necessity of a rigorous personnel selection process. The U.S. Navy's current Aviation Selection Test Battery (ASTB) is predictive of aviation training outcomes. In a recent validation study, it was found that ASTB composite scores had the following values for predicting final grades upon completion of U.S. Navy Primary Flight Training: Academic Qualification Rating $r = .45$ [$p < .001$], and Pilot/Flight Officer Flight Aptitude Rating $r = .35$ ($p < .001$) (ASTB, 2006). Although these correlational values are reasonable, an incentive exists to strive for continued improvement. It is estimated that $6M (million) training dollars per year or more can be saved for each 5 percentage points of variance in training performance accounted for by means of personnel selection (U.S. Navy Flight Surgeon's Manual, 1989).

Among cognitive tests in a wide range of talents, general intelligence/ability accounts for roughly 50% of common variance, while quantitative, spatial, and verbal ability each account for approximately 8%-10% of the remaining common variance (Lubinski, 2004). It follows that if even 1% of common variance can be accounted for via the increased capacity to quantify relevant aptitude and/or ability, the U.S. Navy could save $1M training dollars or more per year. This provides the rationale to investigate additional forms of assessments for U.S. Navy Aviation personnel selection.

Although this work did not go so far as to investigate predictive validity of tasks, its purpose was to show whether the tasks under investigation possess psychometric properties that are desirable for further study. In particular, this work focused on determining the reliability and construct validity of performance-based measures (PBMs) consisting of spatial and psychomotor ability assessments, as these assessments appear to offer an attractive source of new approaches to personnel testing.

Research Goals

This research constituted an initial validation effort with regard to two novel computer-based PBMs: the Block Rotation Task (BRT) and (2-D) Compensatory Tracking (CT). Results from this work may be used to indicate utility of the novel assessments for aviation personnel selection in the U.S. Navy.

The BRT is a derivative of the Shepard-Metzler (1971) mental rotation task. It consists of a set of virtual 3-D blocks: one being the target stimulus and the other being the comparison figure. Participants' goal in this task was to manipulate the comparison figure, in 3 dimensions, into a matching orientation to the target figure as quickly as possible. The CT task is the U.S. Navy's version of a computer-based task that has a strong presence in the literature. This task required participants to keep a cursor in the center of a monitor screen even though the cursor moved in a variable fashion. Participants had limited control of the behavior of the cursor in 2 dimensions.

Both tasks possess relatively short administration times, with no more than 20 minutes' administration time for the BRT, and one minute or less for CT. BRT was hypothesized to make demands on mental rotation (a form of spatial ability), psychomotor ability, and their coordination. CT was thought to make more of a demand on "pure" psychomotor ability and less on spatial processing.

These tasks are being considered for operational use due to the fact that PBMs have historically proven to be good predictors of performance in flight training and that any increase in the predictive power of personnel selection tests yields exponential returns in training dollars. The analog devices that were used for aviation personnel selection were discontinued in the middle of the last century because they were not easily "co-calibrated" and were located in a few, isolated stations, a situation that conflicted with evolving recruiting practices. However, given the ubiquity and power of modern computing devices, the constructs that were previously measured with analog devices can now be measured with "digital" devices.

In essence, this work examined various potential measurement capabilities and features of the BRT. This was achieved primarily via comparison of performance on the BRT with that of other validated performance-based measures. Specifically, the hypotheses of this study were that (more detailed descriptions are contained in Chapter 4 [Results]):

I) BRT correlates with both psychomotor and mental rotation tests while psychomotor and mental rotation tests do not correlate

II) Validated performance tasks express differential predictive models for the BRT in an hierarchical manner

III)	BRT normative performance expresses a linear chronometric quantitative model similar to that found in Shepard-Metzler (1971)

CHAPTER 2: LITERATURE REVIEW

Spatial Ability

Operational Definition

Lohman (2000) indicated that like verbal ability, spatial ability (SpA [distinguished from the oft-used abbreviation for situational awareness, SA]) is a form of general fluid ability (intelligence), expressed in the form of inductive reasoning as spatial processing occurs. As a stand-alone construct, *SpA can be operationally defined as "the ability to generate, retain, retrieve, and transform well-structured visual images whose properties include location, size, distance, direction, separation, connection, shape, pattern, and movement" (Lohman, 1993, p. 1).* Lohman's definition pervades the literature.

Not unlike other measures of cognitive ability, properly quantifying SpA performance includes explaining systematic individual differences that are uniquely spatial, and defining the portion of variation on spatial tasks that is shared with more general abilities (Sternberg, 2000). Sternberg's approach to parsing variability is applied in this research in the sense that shared variance among selected spatial and psychomotor tasks is examined.

Associated Neurological Processes

Neurologically, the processing of spatial information is fundamentally distributed throughout the visual-perceptual system. Spatial processing is traditionally considered to be a right hemispheric phenomenon given the dominance in left hemispheric structures for processing linguistic material.

However, experimental findings have shown that spatial processing, although residing to a significant degree in the right hemisphere, can be expressed in a distributed fashion. After the bifurcation of the early visual structures into ventral and dorsal streams in the lateral geniculate nucleus, some spatial processing occurs in the inferior temporal area, but structures in the posterior parietal area are closely associated with spatial processing proper (Kandel & Wurtz, 2000). Additionally, Kandel, Kupferman, and Iverson (2000) presented evidence that processing spatial material from memory produces activation in the right hippocampal, parahippocampal, and parietal areas. Despite studies that continue to corroborate specialization in the right hemisphere for spatial processing, Levin, Mohamed, and Platek (2005) also implicated the left parahippocampal gyrus, areas of the frontal gyrus, and other frontal and parietal areas in spatial processing.

Regardless of lateralization, all findings show support for a “hard-wired” biological system that is developed specifically for processing spatial information. Evidence for such a system leaves no doubt that a SpA latent construct is valid. In addition to theory driven by cognitive science, this fact provides a rational foundation upon which human performance/ability investigations can be based. However, the question of how to *properly quantify* SpA as a cognitive construct has been a point of interest for many psychometricians and is a primary focus of this research.

High-Level Definitions of SpA Constructs

SpA is an underrepresented component of the human cognitive repertoire in terms of application of psychometric outcomes relative to other factors such as quantitative and verbal abilities.

Nonetheless, scientists have recognized SpA constructs as an empirically central aspect of human intelligence as early as the beginning of the 20th century (Binet & Simon, 1916). Thurstone (1938) proposed that human intelligence is constituted of seven independent factors that he referred to as “primary abilities,” including word fluency, verbal comprehension, spatial visualization, number facility, associative memory, reasoning, and perceptual speed. More recently, Snow (1996) provided a simpler model, with intelligence being constituted by a complexity dimension and three content domains: quantitative/numerical, spatial/mechanical, and verbal/linguistic. In addressing the spatial component of intelligence specifically, Thurstone (1938) recognized that the SpA construct can be further divided into three sub-components: (a) object recognition from different angles, (b) imagining movement/displacement of constituents of a spatial configuration, and (c) determining spatial relationships with respect to one’s body. Subdivision of SpA has also been supported in the recent literature. Sternberg (2000) stated that SpA is not a unitary construct but possesses several subcomponents, each emphasizing different aspects of image generation, storage, retrieval and transformation. Subdivisions of the SpA construct have been well-developed by Carroll (1993) (and subsequently widely adopted by researchers in the field), who showed that the general spatial construct consists of a hierarchy of sub-constructs, in which the performance variance of a generalized construct is split among more specific SPA forms abbreviated here:

- General Spatial Visualization factor: The general spatial visualization factor (Gv or Vz) is at the hierarchy's pinnacle due to relative complexity of processing; it can be measured and quantified by, among others, paper form board, paper folding, and mental rotation tests (e.g., Shepard-Metzler [1971])
- Orientation (SO) factor: Research participants are asked to imagine how an array would appear from a different perspective and then to make a judgment based on the imagined perspective (e.g., an aerial orientation test)
- Rotation (SR) factor: This factor emerges if two or more simple, highly speeded mental rotation tasks are included in a test battery (e.g., flags, embedded figures, mental rotation; Shepard-Metzler [1971] also fits here)

Sternberg (2000) indicated that Vz tests appear to be primarily measures of general intelligence, are secondarily measures of task-specific functions, and thirdly an "undefined" construct that covaries uniquely with other Vz tests. He went on to indicate that SO factors can be distinguished from Vz factors, and that the SR factor emerges if two or more simple, speeded tasks are included in a test battery.

An Ecological Explanation for SpA Constructs

Spatial visualization and mental rotation cognitive functions serve as representations of one's own body in space, as it is understood in terms of its relationship to environmental conditions, including gravity, or the location of other objects in space relative to one another and to the observer's body. The natural condition of orienting one's self with regard to the position of environmental objects has been explicated by Amorim, Isableu and Jarraya (2006).

They indicated that “spatial embodiment,” or giving human characteristics to non-human geometric shapes, helps to improve performance in mental rotation tasks. They attributed this improvement in performance to the general idea that human spatial cognition is action-oriented, and is necessarily a function of taking into account the human body and its spatial and motor representations. Such a conception generally provides for the validity of investigations that consider mental rotation and spatial orientation-types of constructs. From a face-validity perspective, tests that require participants to perform tasks, which appear to require the same resources as real-life spatial activities, simply “fit” the action orientation of SpA. For construct validity, paper and pencil or computer-based (virtual) analogs of real-life conditions that demand those spatial resources as are employed in similar real-life conditions, can be shown to “fit” quantitatively. Perhaps it is this embodiment effect that provides spatial measures with generally strong face validity and exceptionally strong ecological validity as long as the testing task is an accurate reflection of relevant real-world conditions.

Gender Differences

Gender differences are an important aspect of SpA. A majority of modern SpA investigations have gender differences at their root. Historically, findings have shown performance advantages in males with regard to SpA metrics. An example is a recent study (Geary & DeSoto, 2001) which showed significant differences in performance between males and females, with males being “over-represented” at the high end of the performance scale in a battery of SpA tests, and females being “over-represented” at the lower end of the performance scale.

Results also showed that such differences were independent of culture (comparing participants from U.S. and Chinese populations), the implication being that genetic-evolutionary processes lie at the heart of SpA gender differences (Geary & DeSoto). An example of "spin-off" gender difference research is demonstrated in the work of Kass, Ahlers, and Dugger (1998), who showed that gender differences in visual SpA (estimation of orientation angle of a ship viewed through a submarine periscope simulator) could be reduced via training and feedback. Another example of such research is the work of Bowers, Milham, and Price (1998) who compared performance of males and females on different SpA tests and other tasks in an effort infer differences in brain lateralization by gender. Results indicated no systematic SpA differences, and the authors implied that there are no differences in lateralization processes. A further implication of this study has been that the SpA construct has not historically been measured correctly. Bowers et al. indicated that, given the multi-faceted nature of SpA, "inconsistencies" in the literature with regard to gender differences have been found because researchers normally use a single measure of SpA without identifying the SpA subtype it is intended to measure. Such a critique of approaches to measuring SpA is in line with Sternberg's (2000) admonition to accurately measure a construct, but within the context of Bowers et al.'s (1998) work, the admonition also includes identifying the unique variance of SpA subtypes.

Given the two examples presented, gender differences are central to most discussions that involve SpA psychometrics. Research that involves the investigation of gender differences can have implications in the way that SpA constructs are conceived and in the way that they are subsequently measured.

The ultimate outcome of the debate can have repercussions within the realm of validity. For example, Bowers, Milham and Price (1998) investigated the degree to which lateralization of spatial processing, and in effect, gender differences, are artifacts of test selection. Their findings suggest that, prior to administration, tests of spatial ability should be screened for gender-related sensitivity given that completion of spatial ability tasks makes demands on varying levels of skills across measures. This outcome is explicit in Bowers et al.'s (1998) study, because the question of whether a specific measure of SpA, the Guilford-Zimmerman Spatial Orientation test, is an effective SpA measure was addressed in the study. The study confirmed the efficacy of the test.

Quantification of SpA

A consequence of not including SpA in current approaches to measuring human performance, according to Shea, Lubinski, and Benbow (2001), is that modern talent search procedures currently miss approximately 50% of the top 1% in three-dimensional spatial visualization. This finding has implications in various vocational domains in which SpA plays a role, such as surgery, engineering, mathematics, and flying, especially if a goal in such professional communities is to obviate potentially outstanding performers.

Spatial abilities have been measured with performance tests, paper-and-pencil tests, verbal tests, and film or dynamic computer-based tests (Lohman, 1993). Performance tests are among the earliest measures of SpA and have included form board, block manipulation, and paper-folding tasks, for example (Binet & Simon, 1916). According to Lohman (1993), individual differences on most spatial tasks appear to be well accounted for by performance on factors defined by paper-and-pencil tests.

This is reflected by the fact that most spatial tests that are generally administered have taken the paper and pencil format. An example of the paper and pencil approach to measuring SpA is provided by the ASTB's Spatial Apperception Test (SAT). Figure 1 provides an example SAT item. In the SAT, participants are asked to determine the orientation of an airplane based on the rendering of a pilot's view of the horizon. The SAT is very similar to the historically more widely administered Guilford-Zimmerman (1948) Spatial Orientation test.

Verbal tests measure SpA by requiring participants to listen to a problem in which a mental model must be created, and to verbally respond. Although verbal tests of SpA are not often used, they have shown high correlations with other spatial tests and various criterion measures (Ackerman & Kanfer, 1993; Guilford & Lacey, 1947). Verbal measures of SpA often challenge the test taker to solve spatial problems posed in a scenario-based format. For example, Ackerman and Kanfer (1993) provided an example of the Verbal Test of SpA (VTSA). In this test, participants are asked to close their eyes and imagine items described verbally (textually), and they are asked a multiple-choice question. The following is an excerpt from Ackerman and Kanfer's work in which the VTSA was administered:

> It is morning and you are facing east looking at the sunrise. You walk forward for 100 yards, turn left, and after walking another 50 yards you turn about (i.e., turn 180 degrees). In what direction are you now facing? a) North b) South c) East d) West (Ackerman & Kanfer, 1993, p. 431).

Some measures of SpA are more dynamic and attempt to tap into one's ability to adapt to object trajectories and/or arrival times, and often include a psychomotor performance component. An example of such a test is provided in Figure 2, showing a screen shot of the CT task. In this task, performance is measured as the participant's ability to place and maintain the cursor in the center of the screen, using a joystick control. Participants in this task must continuously adapt to random movements in the cursor however, requiring compensation for changing movements in the cursor that are beyond the participant's control.

Such measures, by default, include an element of psychomotor ability given that responses cannot be made without behavioral input of the participant. However, it must be noted that such measures are used to select personnel who possess highly effective combined spatial and psychomotor abilities for very specific tasks, such as military aircraft piloting.

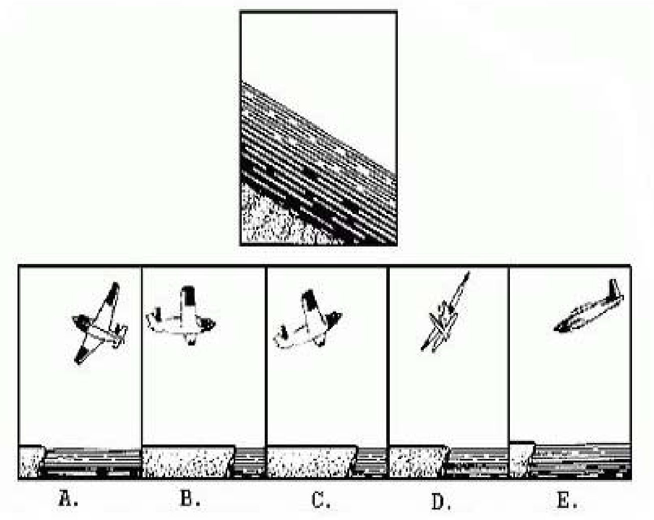


Figure 1. Sample Spatial Apperception item.

In the early, through mid-, 20th century, the use of performance based/SpA tests for military pilot selection proved to be a valid effort in the prediction of primary flight training in the U.S. Navy (Fleishman, 1956). Also, performance based/SpA tests were shown to be good predictors of success in training courses for aircrew positions (Guilford & Lacey, 1947). However, these tests were comprised of full-sized cockpit simulator-like apparati, such as a stick-and-rudder control device or the Ruggles Orientator (Figure 3). As pilot selection testing moved away from military bases and onto college campuses, administering such tests became unfeasible due to the cumbersome nature of the devices and the difficulty of ensuring common calibration among aparati.

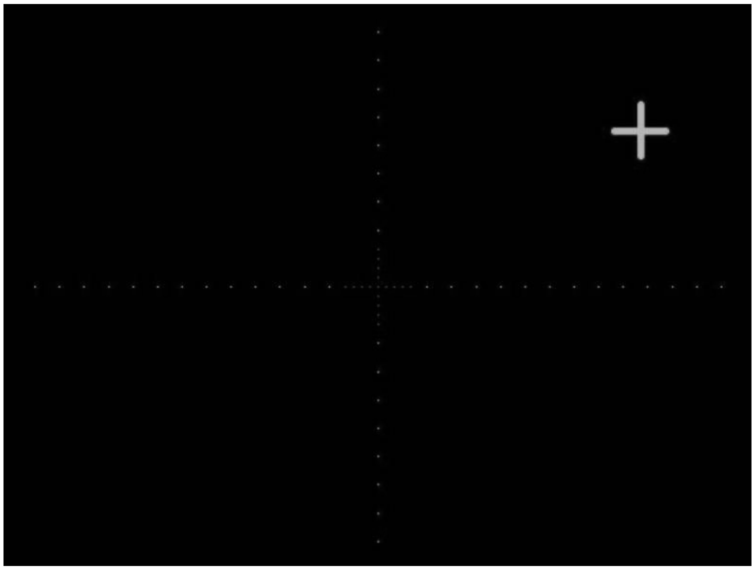


Figure 2. 2-D CT screen shot.

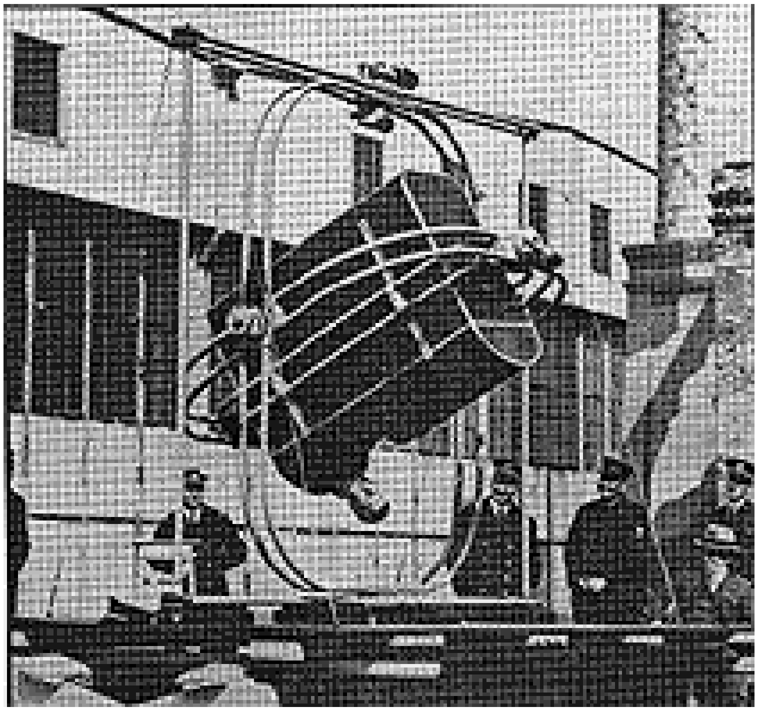


Figure 3. Ruggles Orientator.

Renewed interest in performance-based tests has accompanied advances in computer technology, as tests can now be administered remotely and results can be sent to military administrators via secure network connections. In addition, computer-based tests offer the opportunity to gather both error and latency scores, which can then be combined to predict criterion performances with greater precision than from either measure considered separately (Ackerman & Lohman, 1990). Research has been conducted demonstrating that computerized performance-based tests add a component of prediction to outcome based on ecological validity (e.g., Ackerman, 2001). Ackerman's research supports early work done in the military aviation selection arena. For example, Griffin (1987) showed that performance on computer-based dichotic listening and psychomotor tasks were significantly correlated with U.S. Navy primary flight training grades.

Carretta (1989) showed that computer-based tracking related more strongly with a pass/fail criterion in U.S. Air Force flight training than any other test that was used in a battery of tests used for pilot selection. Delaney's (1990) follow-on study to Griffin's work, but using a much larger sample, showed that computer-based dichotic listening and psychomotor tasks were significantly related to U.S. Navy primary flight training grades. These studies have provided evidence that computerized performance based measures have great utility in accounting for unexplained variance in training outcomes. The study proposed here will continue along that line, but with a focus in the specific area of SpA, with a particular interest in mental rotation.

Mental Rotation

For the purposes of this research, mental rotation is operationally defined as *the ability to rotate mental representations of two-dimensional and three-dimensional objects*. It has been understood that a general SpA construct is constituted of sub-factors, such as those found in Thurstone's divisions of SpA, the first empirical work that focused specifically on mental rotation, as a relatively orthogonal psychometric construct was not conceptualized until the early 1970s (Shepard & Metzler, 1971). An example of the experimental stimulus is provided in Figure 4. Results of their chronometric study (abbreviated in Figure 5) showed that recognition and psychomotor response latency of two-dimensional renderings of three-dimensional shapes were a positive linear function of the angular difference in the portrayed orientations and that this latency did not differ depending on the number of dimensions in which rotations occurred. That is, one-dimensional (1-D) rotation latency is not less or greater than rotation in three dimensions (3-D).

Shepard and Metzler's results also showed that the slope of the obtained functions represented an average rate at which the stimuli were mentally rotated of approximately 60° per second, including the time it took to execute psychomotor responses.

It is theorized that specific events occur during mental rotation. According to Johnson (1990), mental rotation can be separated into the following stages:

- Creation of a mental image of an object
- Rotation of an object mentally until a comparison can be made
- Make the comparison
- Decide if the objects are the same or not
- Report the decision

It is possible that increased angular disparity between target and comparison stimuli demands more of decision-making processes and results in increased response latency.

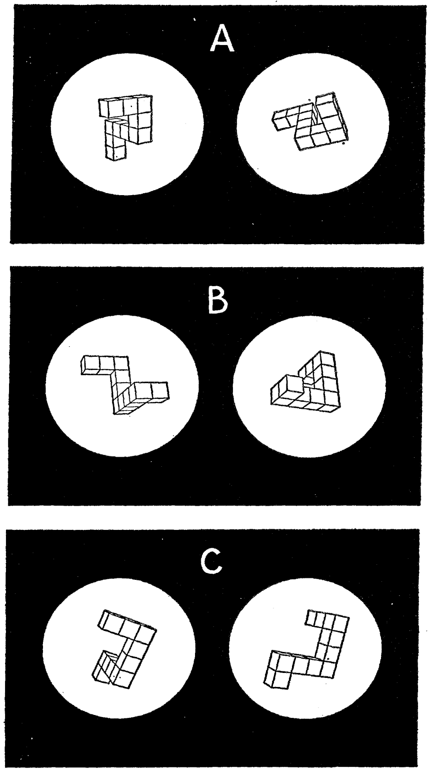


Fig. 1. Examples of pairs of perspective line drawings presented to the subjects. (A) A "same" pair, which differs by an 80° rotation in the picture plane; (B) a "same" pair, which differs by an 80° rotation in depth; and (C) a "different" pair, which cannot be brought into congruence by *any* rotation.

Figure 4. Mental rotation stimuli used by Shepard and Metzler (1971).

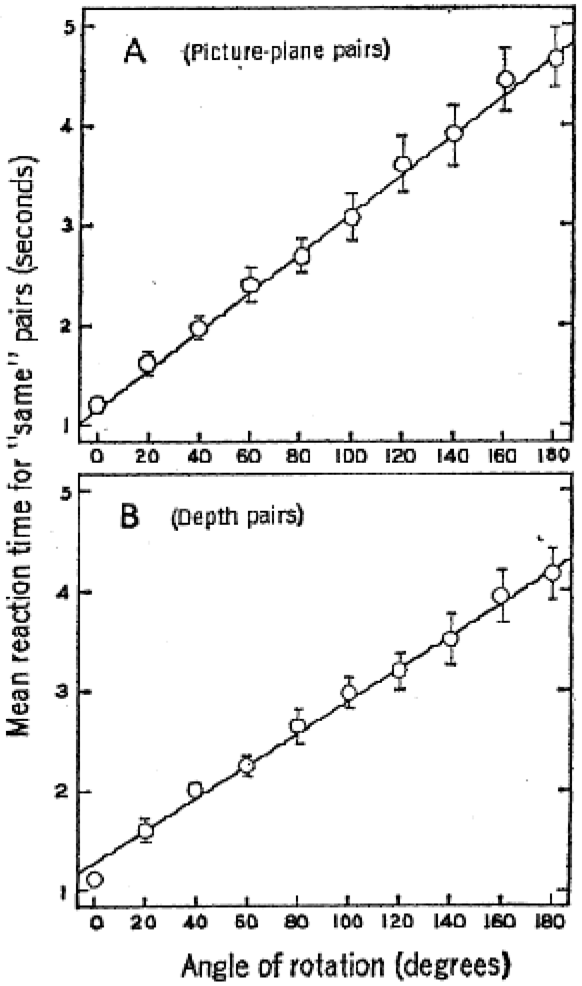


Figure 5. Chronometric results of Shepard-Metzler's (1971) experimental protocol

Psychomotor Ability

Operational Definition

Psychomotor ability is defined as that which reflects the "*capabilities of the motor system to plan, coordinate and execute movements*" (Ghez & Krakauer, 2000, p. 653). Guilford (1958) examined psychomotor ability and defined numerous aspects of psychomotor functioning and indicated that functioning could be classified according to the following kinds of abilities: strength, impulsion, speed, precision, coordination, and flexibility. Of the aspects of functioning presented, those most pertinent to the current study include precision (of directed movements) and coordination (of different limb movements).

Associated Neurological Processes

Motor processing is the reverse of the sequence in the sensory system (Saper, Iverson, & Frackowiak, 2000). Motor planning commences with a "general outline" of intended behavior and is transformed into actual responses in motor pathways. Patterns of frontal neurons firings constitute the source of individual and complex motor actions. The motor pathways leaving the cerebral cortex have their origin in the primary motor cortex of the precentral gyrus. Neurons in the premotor cortex are associated with planning of movements and receive inputs from motor centers in the thalamus, primary somatosensory cortex and the prefrontal association cortex. Although the premotor area sends projections to other areas of the brain, of great import are projections to the primary motor cortex (Saper et al., 2000) and cerebellum. Essentially, the primary motor cortex sends projections down the spinal cord to synapse with motor neurons that connect to muscles, with the cerebellum being central in motor control.

Generally, psychomotor movements are controlled by two kinds of systems: feedback control and feedforward control (Ghez & Krakauer, 2000). In feedback control, signals from sensory organs are compared to a reference signal by a "comparator." Error, if any, is corrected by a change in output by a controlling mechanism. Feedforward control is determined by information acquired before feedback sensors are activated. Accuracy in movement using feedforward control requires prior knowledge of stimulus characteristics so that appropriate movements can take place. Both of these systems will play a role in the psychomotor tasks that are being employed in the current study.

Application of Psychomotor Data

An example of how psychomotor data can be applied in personnel selection is provided by the work of Johnston and Catano (2002). They examined the predictive and incremental validity of three psychomotor ability measures: manual dexterity, finger dexterity, and motor coordination. The researchers combined these measures with cognitive assessments and found that the three psychomotor measures increased validity for selected technical and mechanical occupations. Their findings suggest that PBMs such as psychomotor tasks can improve predictions beyond the sole use of cognitive measures when job analyses reveal that a given assessment is relevant.

Although the aim of this particular study is not to demonstrate such ecological validity for the tasks that are being examined, this work should pave the way for follow-on work that explores relationships between the experimental tasks that are being examined here and performance in U.S. Navy flight training. Additionally, examination of performance on the experimental tasks should not be limited to flight training, but other applications (e.g., medical applications) should also be explored.

Computer-Based Medium

In support of the notion that spatial constructs can be measured digitally, Pelligrino et al. (1987) demonstrated a computer-based battery that assessed SpA factors. These researchers also found that an integrated software package constituted a valid performance measurement system for the identified constructs.

When a measure has been traditionally administered using a paper-and-pencil medium, it is necessary to indicate that the validity of the measure is not lost when a change in administration medium has taken place. Larson (1996) indicated that paper-and-pencil SpA tests can measure the same construct as computer-based tests. Finally, the evidence associating SpA with general ability *g* also shows that SpA cross-validates with other cognitive constructs inherent to aptitude tests that are administered for aviation selection purposes.

Current technology enables us with the ability to present data to users and compile information created by users in rapid fashion and with great accuracy. Computer-based testing media permit the presentation and collection of both static and dynamic data, making the inclusion of performance based measures possible at sites that are remote to central testing organizations. For example, at selected administration sites the U.S. Navy is offering a web- and computer-based version of the ASTB, referred to as the Automated Pilot Exam (APEX). This version of the ASTB allows for administration in distributed areas, with a centrally-located secure server and scoring center (Naval Operational Medicine Institute, 2004).

Notwithstanding the reduced time in test administration, scoring and reporting, platforms such as APEX present an ideal means by which performance-based tests can be administered. However, it must be noted that APEX does not include any form of performance assessment.

Summary

Spatial and psychomotor ability are constructs of interest in Naval Aviation personnel selection given that such abilities are critical for the safe operation of aircraft and for accurate navigation. Additionally, such latent constructs could prove to be the source of untapped explained variability with regard to human performance measurement, which if tapped, could yield significant savings in terms of training dollars.

The purpose of this work was to determine if the selected experimental tasks (i.e., BRT and CT) are reliable and valid indicators of performance within the identified domains. Specifically, a goal was to determine if the BRT, as a measure of both mental rotation and psychomotor ability, correlated with CT (a measure of psychomotor ability), and previously validated SpA measures. Reliability coefficients and correlations among experimental spatial and psychomotor ability measures were also a significant area of interest. Other goals were to demonstrate the degree to which the BRT correlated with measures of Carroll’s (1993) SpA sub-constrcuts and to provide quantitative human performance models of the BRT and the CT.

CHAPTER 3: METHOD

Participants

Students from the University of Central Florida (UCF) were solicited to volunteer for this study via the UCF Psychology Department in accordance with Federal law and with the requirements of the University of Central Florida's (UCF) Institutional Review Board (IRB) (see Appendix A). All participants were briefed on the protocol, informed of the voluntary nature of participation and asked to sign an Informed Consent form (Appendix A). All participants were assigned a randomly generated participant number to protect privacy. No personal identifiers were used on materials containing participant data. All data were maintained in a secure, locked location. Instructors provided extra credit in Psychology courses for incentive for students to participate.

Although power analyses were conducted for each particular statistical procedure that was used to address the hypotheses of this study, a rough idea of the number of participants that would be required for this study was provided by Cohen (1992), who indicated that $N = 42$ is sufficient for power = .80 and $\alpha = .05$ to detect large effects of 5 variables using multiple regression and correlation analyses.

Research Design

A primary focus of this work was to determine if BRT performance was related to that of other validated SpA tasks. To accomplish this, the approach taken was to compare of scores on the BRT (number correct), CT (mean deviation from target), Spatial Apperception (number correct), Mental Rotation (number correct), and Manikin (number correct) tasks. Other hypotheses could also be addressed by using this repeated-measures design, where each participant was asked to complete a battery of tests.

Materials

All assessments contained instructions and sample problems. Descriptions of the tasks that were used include:

Pre-Simulation Questionnaire (adapted from Qu, 2003) (Appendix B)

This survey solicited demographic information from the participant. It was intended to be used as a resource for additional information about participants should a confound have impacted the results. Approximate administration time: 5 minutes.

Spatial Apperception (SAT)

Paper-and pencil administered test in which participants selected an outside static view of ground-aircraft relation based on static presentation of horizon as seen from the cockpit. Correct identification of targets required spatial relations and orienting capabilities. Performance was measured in terms of the total number of correct items. Approximate administration time: 25 items in 10 minutes.

Block Rotation Task (BRT) (Figure 6)

The BRT is an automated variant of the Shepard and Metzler (1971) mental rotation task. On the left side of the computer screen, the 3-D BRT presented a static 3-D figure constructed from blocks, referred to as the goal figure. An identical comparison figure was presented on the right side of the screen, but was always in a different orientation in terms of its rotation along any one of the x, y, and z axes individually or in combination. The objective was to use a joystick and throttle to rotate the comparison figure into exact alignment with the goal figure. After administering a practice session, the computer allowed one minute to solve each of 24 problems and recorded the time required (up to one minute) to solve each problem.

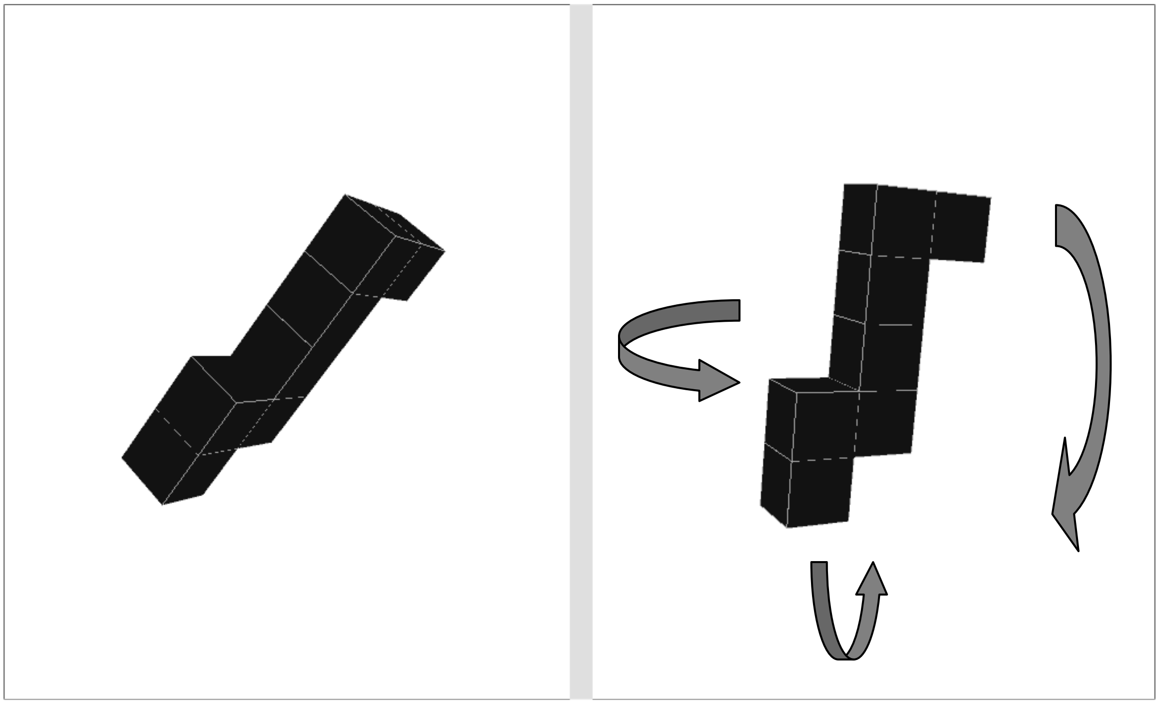


Figure 6. Sample Block Rotation Task item

The computer also recorded error (deviation of the comparison figure from the target figure) in the subject's solution in each of the three axes. Error was measured on a 100-point scale, where 1.00 is the greatest distance between the rotational axis of the movable figure and that of the target figure and 0.00 is a perfect match between the two. Directionality of separation between axes was indicated by positive and negative values.

If the participant did not solve the problem in the allotted time (30 seconds), the computer automatically moved on to the next problem and scored the missed problem as an error. Otherwise, participants could elect to move on to the next problem if he/she was satisfied with the solution. Performance was measured in terms of the total number of correct items as well as other descriptive data (e.g., combined spatial and psychomotor human performance modeling), that were explored for the first time ever here in this study. Sampling rate was 500 ms. Approximate administration time: 25 minutes.

Vandenberg Mental Rotation Test (MRT)

This paper-and pencil administered test's 20 items were organized into five sets of four. Items consisted of a criterion figure, two correct alternatives and two incorrect distracters. Correct alternatives were identical to the target shape. In half of the items, distracters were rotated mirror images of the criterion while in the other half, distracters were rotated images of criteria in other items. Items were scored correctly only if both correct alternatives are identified and were considered incorrect if both correct alternatives are not identified. Performance was measured in terms of the total number of correct items. Approximate administration time: 10 minutes.

Manikin Test (Lane & Kennedy, 1990)

In this computer-administered test, participants had to determine which hand, right or left, was holding an object that matched the object on which the manikin was standing. The manikin was positioned standing upright facing either toward or away from the participant, or upside down, also facing toward or away from the subject. The manikin's position was distinguished by characteristics such as facial features and clothing. Responses were made by pressing an arrow key, where the arrow pointing left is indicative of the object being held in the left hand and the right arrow represents the object being held in the right hand. Performance was measured in terms of the total number of correct items. Approximate administration time: 1.5 minutes.

2-D Compensatory Tracking

Computer-administered task in which participants were asked to keep a randomly drifting cursor in the marked center of a computer display screen. Scoring consisted of a 100-point scale using a traditional Cartesian coordinate system (that includes negative values), where 1.00 is the greatest distance between the crosshairs and the center of the screen and 0.00 is the least distance between the two. Sampling rate was 100 ms. These data were transformed into root mean square error (RMSE) data for use in statistical analyses. Approximate administration time: 5 minutes.

Procedure

Participants were solicited via the required UCF on-campus recruitment procedure (Sona Systems), were briefed and asked to read and sign the consent form. and administered the pre-simulation survey. Experimental tasks were administered to participants using laptop computers (Dell Latitude D620). Specific procedures that were used for the experiment are listed. All participants were briefed on the protocol and were asked to read and sign the UCF-approved Informed Consent Form (Appendix A).

The BRT, Spatial Apperception Test (SAT), Vandenberg Mental Rotation Test (MRT), Manikin (M) test and CT were administered in a generally counterbalanced fashion. However, due to constraints inherent to the novel software, BRT and 2-D CT were administered in the same order. After being assigned a randomly-generated participant number to protect privacy, participants were administered the pre-simulation questionnaire and the spatial abilities battery (~60 minute administration time total). Upon task completion, all participants were thanked for their time and effort and were solicited for questions or comments that they may have had with regard to the protocol.

CHAPTER 4: RESULTS

Introductory Remarks

Data analysis procedures used in this study consisted primarily of descriptive statistics, multiple regression/correlation and ANOVA. Other tests were conducted on an as-needed basis. Standard level of significance for statistical tests was set at the customary $\alpha = .05$. Statistical tests were run using SPSS 10.0. Power analyses were conducted using G-Power 2.0 (Buchner, Erdfelder, & Faul, 2001).

After data screening and reliability tests were completed, gender differences in performance were examined to determine if performance on spatial ability concords with gender differences that pervade the literature. This analysis was conducted simply to demonstrate the validity of tests in the sense that gender-related findings in this study are similar to those found in other studies. Second, a correlational analysis was conducted to determine the independence or relatedness of task performance to infer corresponding latent construct behavior. This analysis was conducted to show convergence and divergence of SpA and psychomotor constructs. Regression analyses were run to determine: (1) the existence and form of models consisting of validated measures in their prediction of BRT performance; and (2) the existence and form of models with gender imposed as a factor in the prediction of BRT performance. (1) was conducted to determine which of Carroll's (1993) SpA constructs explained BRT performance given that the validated measures used in this study can be used to represent Carroll's (1993) SpA sub-constructs. (2) was conducted to garner evidence in support of or argumeant against the notion that males and females differ in strategic approaches to spatial problem solving.

Finally, normative data and trends with regard to performance on experimental tasks were examined. These final analyses explicate, at a fundamental level, the reliability of BRT and CT measures.

Table 1

Constructs, Related Validated Tasks and Experimental Tasks

Construct	Validated Task
Visualization*	Vandenburg Mental Rotation
Orientation*	Spatial Apperception
Rotation*	Vandenburg/Manikin
Psychomotor	2-D Compensatory Tracking

*Note: From Carroll's (1993) explication of SpA constructs.

Data Screening

Descriptive Statistics

Some basic demographic and performance descriptive statistics are presented in Table 2. Table 3 shows handedness distributions across gender, where data in both samples approximate the population norm, where it is generally understood that 90% of the population is right-hand dominant.

Data from 73 pseudo-randomly (solicitation was completed via the UCF Psychology Department) solicited participants were collected, consisting of n = 52 (71%) females and n = 21 (29%) males from UCF's undergraduate student population. The male-female proportion disparity may be explained by the fact that there are generally more females than males enrolled in universities across the U.S.

Further, the disproportion is perhaps exacerbated by the fact that there are generally more females than males enrolled in Psychology courses.

Most participants completed all tests. However, some participants did not complete all assessments due to apparently random errors related to understanding or following instructions or due to apparently random errors related to computer operation. However, as reflected by descriptive statistics that accompany each particular analysis in this study, such discrepancies are minimal and do not appear to pose a threat to the validity of these results.

Table 2

General Descriptive Statistics

Variable	N	Min.-Max.	Mean (M)	Standard Deviation (SD)
Females	52	---	---	---
Males	21	---	---	---
Age	73	18-32	19.44	2.30
Manikin Total Correct	69	13-67	40.53	11.01
CT Mean Deviation	73	.17-.81	.39	0.13
CT Standard Deviation	73	.14-.38	.26	0.58
BRT Total Correct	73	0-19	3.48	4.27
SAT Total Correct	73	0-22	10.23	4.84
MRT Total Correct	72	0-20	9.04	5.47

Table 3

Participant Gender-Handedness Demographics

Male		Female	
Left-Handed	Right-Handed	Left-Handed	Right-Handed
2 (10%)	19 (90%)	7 (13%)	45 (87%)

BRT Data Screening and Cleaning

If one wished to examine BRT performance only in terms of the total number of correct items, one would be at a loss for the richness of data that this assessment can provide. The total number of correct items on any validated task can provide valuable information for anyone who wishes to have “quick and dirty” human performance information, and the BRT may well provide that. However, given the complexity of the interaction of *combining* human performance abilities (and in the case of the BRT this appears to be spatial ability and psychomotor performance) a tool that measures such interactions is bound, perhaps even required, to produce data that approach the complexity of the behavior itself. Such is the case with the BRT. The aims of this study required the use of the total number of correct items for correlational analysis as well the use of collective performance data to produce a “normative” human performance model for 3-D processing and associated psychomotor behavior. Total number of correct BRT responses is uncomplicated to derive. However, the BRT software produces an enormous amount of data per participant per testing session. For example, one item consists of approximately 60 data sets if a participant is engaged in the item for the entire 30 seconds.

Multiplied by 28 (the total number of items per test session), this provides the investigator with almost 1700 sets of raw performance data per participant per testing session if each participant takes the full 30 seconds to complete each item.

Graphs representing the completion of a single BRT item follow, illustrating the process used to screen BRT data for normative analysis. Graph lines represent error between the BRT axes (3 dimensions) as well as an additional line representing the mean.

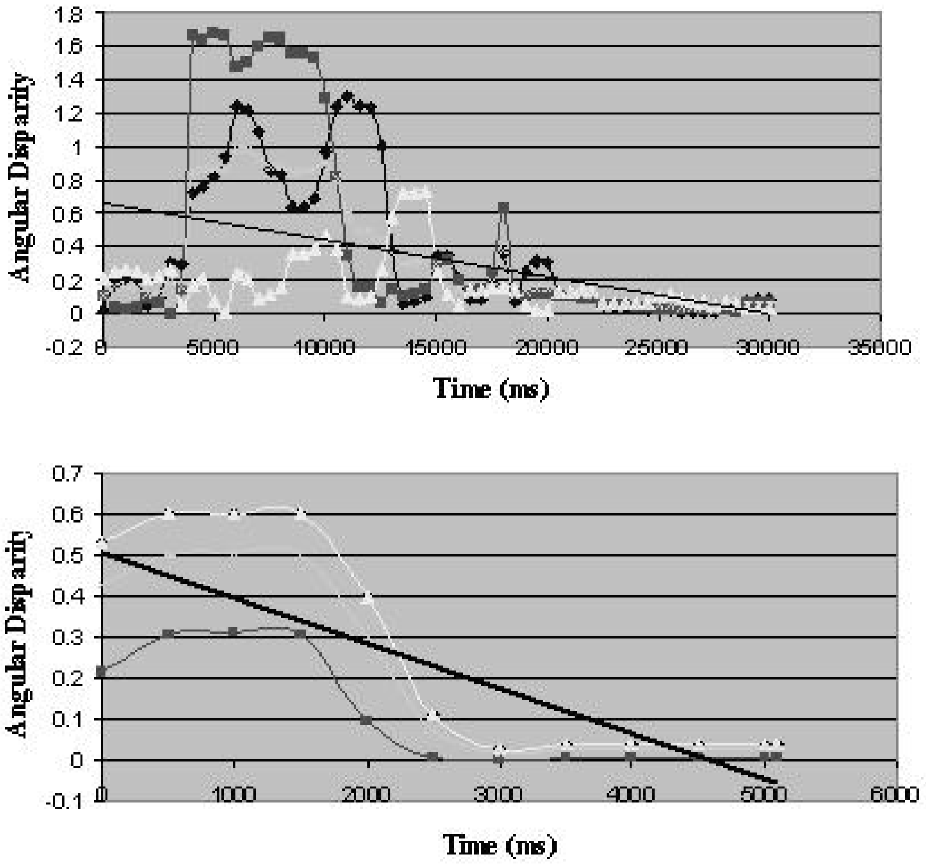


Figure 7. Ideal Performance. The graph on the top characterizes an ideal and most frequently encountered form of BRT error reduction, with an initial perturbation and a gradual reduction of error across all three axes. On the bottom, a characteristic reduction is shown but completed in relatively short order (only 5 seconds to completion).

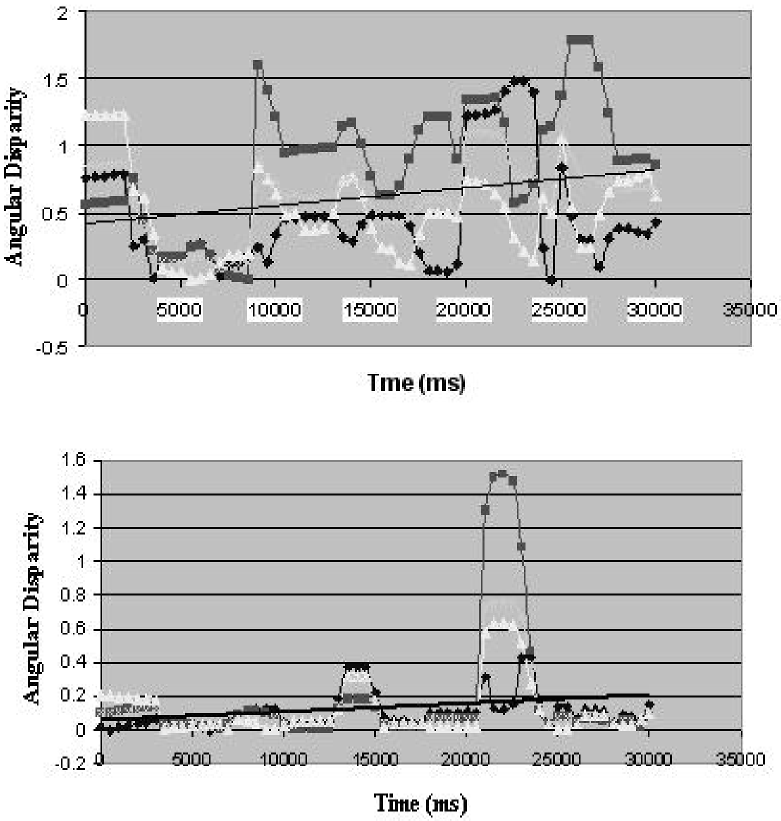


Figure 8. Less Than Ideal, Corrective. Where Figure 7 shows the most common representation of error reduction, these graphs demonstrate the variability with which BRT items can be corrected. The characteristic initial perturbation in Figure 7 is absent, but there is a gradual error reduction as time advances. Evidenced by the normative results, these kinds of data dilute what could be a more linear relationship between time and error reduction. However, given that behaviors resulting in "atypical" data such as these resulted in successful error reduction, the data were entered into the BRT normative performance model.

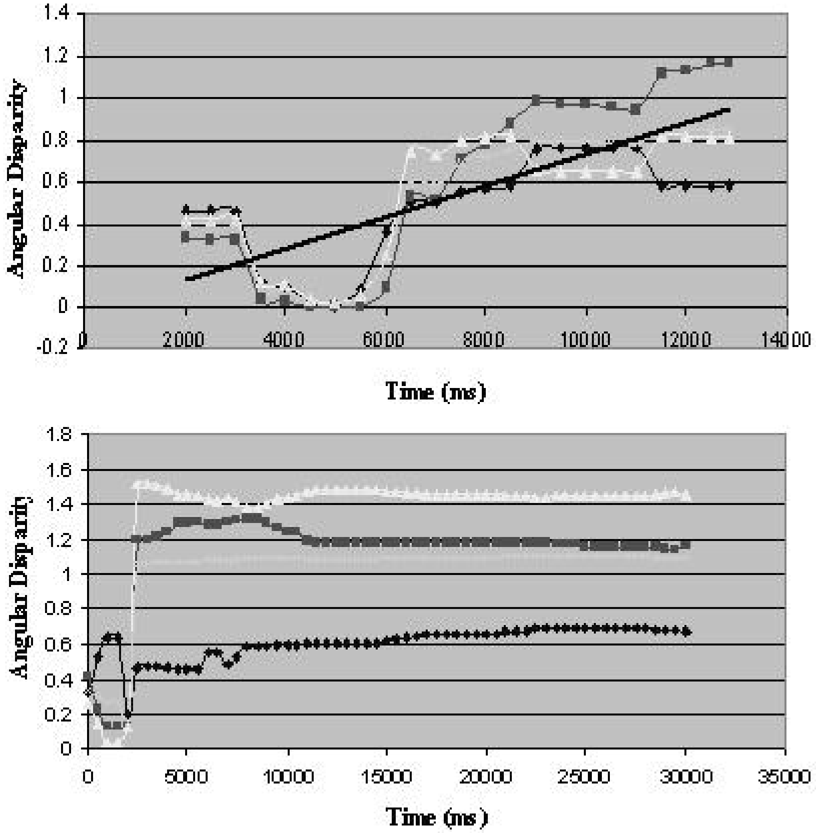


Figure 9. Non-Corrective. Juxtaposed to the examples of error reduction provided in Figures 7 and 8, these graphs show the absence of error reduction altogether. The graph on the top shows that the participant manipulated computer peripherals but was unsuccessful in any error reduction prior to cessation of problem-solving. The graph on the bottom shows no manipulation of computer peripherals indicating the complete absence of an attempt at problem-solving. Data such as these were not included in the normative analysis given that the goal of the normative analysis was to model human performance that reduces error between block stimuli.

Assessment Reliability

Any inference based upon measures made by an unreliable tool is invalid. To address this issue, reliability assessments were conducted on the performance assessments that were used in this study where possible. If a reliability analysis was not conducted for a particular assessment, the raw data were such that it was not possible to properly conduct the analysis. In cases where a reliability analysis was not conducted, reliability findings are cited or are inferred from the raw data to the extent possible.

Reliability estimates for the SAT and MRT in this study were made by transforming correct and incorrect item responses into "dummy variables" (1 = correct and 0 = incorrect). Cronbach's alpha and split-half analyses were conducted for both assessments with results displayed in Table 4. Findings indicate marginal to adequate reliability coefficients for both the SAT and MRT. Lane and Kennedy (1990) indicate that an unreliable test is one with inter-trial correlations of about .70 due to too much error measurement to be useful in repeated measures designs. It notable that the SAT's reliability coefficients are not higher. Assessments in this study were administered to participants only once, so any notion of measurement stability could not be addressed. However, a repeated measures approach was emulated by conducting the specified reliability estimate procedures. Further, although this study does not have a repeated measures component, it is still desirable to conduct this work knowing that results are generalizable and replicable.

A reliability estimate for the Manikin test was not conducted because the software does not make available data for each item response. This is also the case for the BRT in terms of providing data for correct vs. incorrect data, but the richness of the raw data from the BRT can have a compensatory effect. Although reliability data were not available for the Manikin task, Lane and Kennedy (1990) provide some reliability information. They indicate that the Manikin task possesses a “reliability efficiency” estimate of .91, where reliability efficiency is defined as the (normalized) largest reliability likely to be encountered in practical applications.

Reliability data for the CT and BRT assessments were not conducted and perhaps will be conducted in a future study. However, the reliability of these tests may be inferred from their intercorrelations with validated tasks and via inspection of the descriptive models that will be provided when Hypothesis III is addressed.

Table 4

Reliability Analysis for SAT and MRT

Assessment	*N* of Cases	*N* of Variables	*M*	Variance	*SD*	Alpha	Split-Half*
SAT	63	25	10.68	23.16	4.81	.79	.65
MRT	65	20	9.60	28.28	5.32	.88	.72

*Note: Between-forms correlation.

Approach and Analysis

Gender Differences

Tables 5 and 6 show significant gender differences in all measurement domains with the exception of CT Standard Deviation, with males having demonstrated superior performance. This finding concords with most extant literature that examines gender differences with regard to spatial processing and serves as one aspect of the set of indicators in this study that the measures used here are valid SpA assessments. However, a post-hoc power analysis indicated that this test possesses a statistical power of about .56. In order to achieve the desired traditional power of .80, approximately 130 are needed.

Results as they relate to the hypotheses of this study are provided.

Hypothesis I

In line with the expectation that the BRT is a measure of both psychomotor ability and mental rotation, performance on both the psychomotor task (CT) and the mental rotation tasks (Spatial Apperception, Mental Rotation, and Manikin) were expected to positively correlate with the BRT. To demonstrate that the BRT is a measure of two independent constructs, the psychomotor task (CT) was expected to be uncorrelated with any SpA measure. Table 7 shows the correlation matrix supporting the hypothesis that the BRT measures both spatial and psychomotor abilities. However, given that the psychomotor task correlates with all SpA measures, independence of psychomotor (via 2D CT) and SpA constructs was not demonstrated using the performance assessments of this study. Power analysis for an *r* effect size of .30 with a sample size of 73 indicated β at 0.84.

Table 5

Analysis of Variance Descriptive Data

Assessment	Gender	*N*	*M*	*SD*	*Minimum*	*Maximum*
Manikin Total Correct						
	Female	49	38.69	10.23	13.00	61.00
	Male	20	45.30	11.93	27.00	67.00
CT Mean Deviation						
	Female	52	0.41	0.13	0.20	0.81
	Male	21	0.33	0.12	0.17	0.63
CT Standard Deviation						
	Female	52	0.27	0.05	0.16	0.38
	Male	21	0.24	0.07	0.14	0.37
BRT Total Correct						
	Female	52	2.21	2.84	0.00	12.00
	Male	21	6.62	5.54	0.00	19.00
SAT Total Correct						
	Female	52	9.25	3.91	0.00	20.00
	Male	21	12.52	6.18	3.00	22.00
MRT Total Correct						
	Female	51	7.02	4.62	0.00	17.00
	Male	21	13.67	4.43	5.00	20.00

Table 6

Analysis of Variance for Gender Differences

Source	Groupings	Sum of Squares	*df*	Mean Square	*F*	*p*
Manikin Total Correct	Between	619.83	1	619.83	5.38	< .05
	Within	7726.60	67	115.32		
	Total	8346.43	68			
CT Mean Deviation	Between	0.09	1	0.09	5.91	< .05
	Within	1.07	71	0.02		
	Total	1.16	72			
CT Standard Deviation	Between	0.007	1	0.007	2.12	> .05
	Within	0.23	71	0.003		
	Total	0.24	72			
BRT Total Correct	Between	290.59	1	290.59	20.16	< .05
	Within	1023.63	71	14.42		
	Total	1314.22	72			
SAT Total Correct	Between	160.33	1	160.33	7.39	< .05
	Within	1540.99	71	21.70		
	Total	1701.32	72			
MRT Total Correct	Between	657.23	1	657.23	31.48	< .05
	Within	1461.65	70	20.88		
	Total	2118.88	71			

Table 7

Human Performance Correlation Matrix

Ability Measure	1	2	3	4	5	6	7
1. Manikin Total Correct	--	-.21*	-.11	.46*	.32*	.35*	-.08
2. CT Mean Deviation	--	--	.76*	-.31*	-.26*	-.24*	.11
3. CT Standard Deviation	--	--	--	-.18	-.17	-.17	.02
4. BRT Total Correct	--	--	--	--	.54*	.50*	.22*
5. SAT Total Correct	--	--	--	--	--	.47*	.03
6. MRT Total Correct	--	--	--	--	--	--	.02
7. BRT Mean Completion Time	--	--	--	--	--	--	--

*Note: $p < .05$, one-tailed.

Hypothesis II

Not unlike Hypothesis I, it was expected in Hypothesis II that performance on validated tasks would be predictive of performance on the BRT. The correlation matrix in Table 7 establishes that performance on validated tasks is related to that of BRT performance. In the matter of prediction, stepwise regression shows, in Table 8, the degree to which the 3 "best" models predict variance in BRT performance. With an effect size of f2 = 0.15 and sample size of 73, power analysis showed that multiple regression indicated the following β values for (a) one predictor: 0.90, (b) two predictors: 0.83, and (c) three predictors: 0.78.

Table 8

Stepwise Regression Results

Model	R^2	R^2 Change	Change in *p*
1	.32	---	---
2	.40	.08	.005
3	.44	.04	.04

Notes: All models are significant predictors of BRT performance
Model 1 predictor: SAT Total Correct
Model 2 predictors: SAT Total Correct + Manikin Total Correct
Model 3 predictors: SAT Total Correct + Manikin Total Correct + MRT Total Correct

The regression results with regard to Hypothesis II show that the single best predictor of BRT performance is SAT performance. The model with the next highest predictive power is SAT performance combined with that of Manikin performance, providing an additional 8% of explained BRT performance variance. Finally, MRT added to the Model 2 provides an additional 4% explanation in BRT performance.

Gender differences with regard to predictive models were examined using stepwise regression. Results are shown in Tables 9 and 10.

Table 9

Female Stepwise Regression Results

Model	R^2
1	.30

Notes: Model is a significant predictor.
Model predictors: MRT Number Correct + Mean BRT Completion Time + CT RMSE + Manikin Number Correct + SAT Number Correct.

Table 10

Male Stepwise Regression Results

Model	R^2	R^2 Change	Change in *p*
1	.49	---	---
2	.59	.10	-.02

Notes: Models are significant predictors.
Model 1 predictor: SAT Number Correct.
Model 2 predictors: MRT Number Correct + Mean BRT Completion Time + CT RMSE + Manikin Number Correct + SAT Number Correct.

Comparison of models shows that a single model consisting of all tests used in the experiment predicts BRT performance for females and that two models, one consisting solely of SAT (spatial rotation) performance and the other consisting of all tests used in the experiment, predicts BRT performance in males. Results show that the predictor models are stronger for males than for females (R^2 = .49 and .59 vs. R^2 = .30, respectively).

It is notable that in prediction of male performance, Model #1, consisting only of SAT performance, is a relatively high predictor given that the entire set of tests used in the experiment, constituting Model #2, increases R^2 by only .10.

To examine the potential confound effect that time to complete BRT items could have had on these results an hierarchical regression was performed. The results of this analysis are presented in Table 9. Power analysis for the last regression procedure holds true for the current one.

Table 11

Hierarchical Regression Results

Model	R^2	R^2 Change	Change in p
1	.45	---	---
2	.49	.04	.04

Notes: Both models are significant predictors of BRT performance
Model 1 predictors: CT + MRT Total Correct + SAT Total Correct + Manikin Total Correct
Model 2 predictors: CT + MRT Total Correct + SAT Total Correct + Manikin Total Correct + Mean BRT Item Completion Time

Hypothesis III

It was expected that normative data for both the BRT and the Compensatory CT task could be derived from the data. With regard to the BRT, it was expected that a chronometric linear performance model could be derived from the BRT data similar to that of Shepard-Metzler (1971), where the behavior of solving a spatial-psychomotor problem can be chronometrically defined. The model in the current investigation was found to be significant in terms of error reduction with regard to time. The model constrained to linear form is shown in Figure 10.

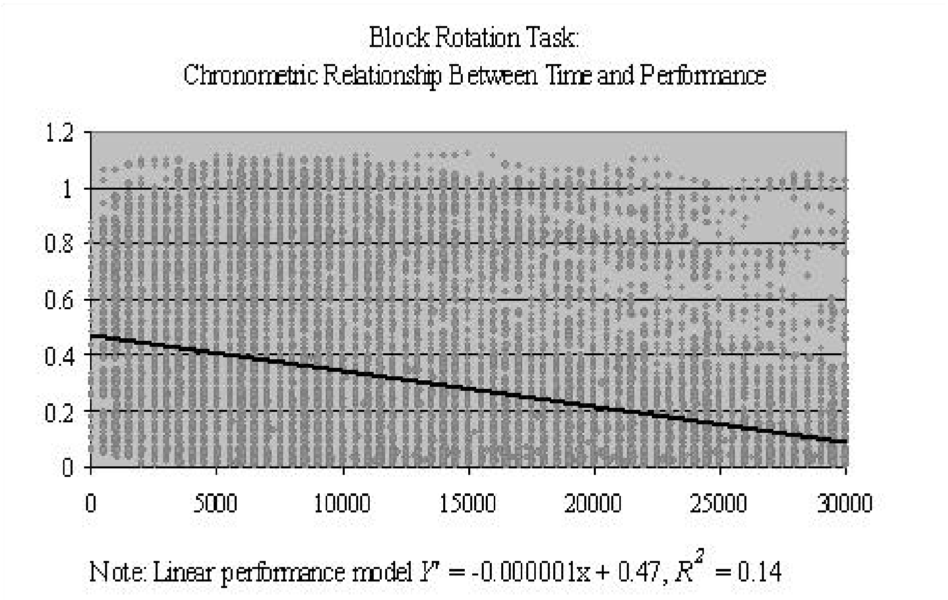


Figure 10. Chronometric relationship between time and error reduction on the BRT ($p < .05$). Data are shown for all participants and all screened test items.

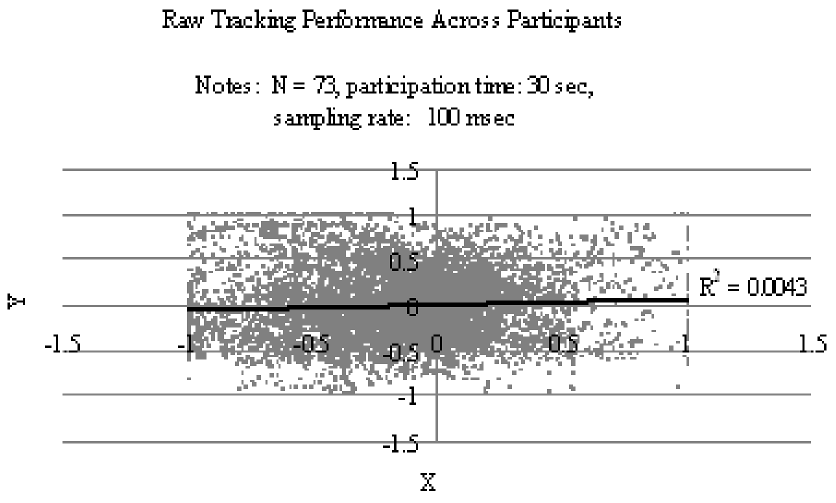


Figure 11. CT Performance Data. CT data across all participants are shown.

Figure 11 shows raw CT performance data for all participants. The mean location along the x axis with regard to cursor location across all participants was –0.11 (*SD* = 0.32). The corresponding value for the y axis was –0.007 (*SD* = 0.33). These data indicate that participants tended to place the cursor slightly left and slightly below the target. Also notable is the flat trend line and associated non-significant R^2 value indicating that the data points appear virtually at random on the graph. This result suggests that data are dispersed equitably among all quadrants and that CT is a fair 2-D CT performance measurement tool.

CHAPTER 5: DISCUSSION AND CONCLUSION

Initial Remarks

This study sought to identify possible BRT latent variables, to determine predictability of BRT performance, and to provide descriptive data for BRT and CT tasks. Those goals were achieved. However before getting into a discussion of the results of this study it is necessary to point out a few items that will help to place the results in perspective. First, although a great effort was made to screen data for the different analyses that were presented, more could be done to produce a cleaner human performance model with regard to the BRT. There was so much variability in the manner in which successful problem navigation was accomplished (as evidenced in Figure 7 and Figure 8) that valuable information would have been lost if models not conforming to the "ideal" expression had been not included in the analyses. However it is possible that the successful completion models that did not conform to the "ideal" form can show a number of possible alternatives to problem solving in the BRT context for future study, should it be conducted. Further, the different approaches to problem solving, expressed in "other than ideal" forms may be a function of item difficulty, an area of interest for item analysis.

Assessments that have been previously validated (MRT, SAT and Manikin) have either shown suitable reliability coefficients in the current study or have shown acceptable reliability in the literature. The reliability of the BRT and CT tasks were not examined in this study with a formal test. This was not completed, in part, due to the unavailability of data with which to conduct such analyses.

However, the model presented in Figure 10 is constituted of well over 100,000 individual data points (Sampling Rate (60/item) x N Items x N Participants), and Figure 11 is constituted of approximately 22,000 individual data points Sampling Rate (300/participant) x N Participants). Until suitable stability and reliability analyses can be conducted on these performance assessments, it is not unreasonable to presume an acceptable "working" reliability for the purposes of this investigation based on intercorrelations of these assessments with previously validated tasks and the appearance of the data that has been shown.

Hypothesis I

Hypothesis I was not fully supported by the results. Although strong correlations existed between the BRT and representatives from SpA and psychomotor domains, SpA and psychomotor domains were not demonstrated to be independent from one another, indicating that these domains to not truly represent what they were presumed to represent. It is possible that the CT task was not independent enough from spatial processing to be considered an observable psychomotor task. This confound may be explained by the fact that the CT task, or the nature of the task, required participants to determine spatial relationships in order to maintain the movable cursor over the target cursor.

However, it is fitting that the BRT correlated with the SAT and MRT to a degree higher than that of the BRT's correlation with CT and Manikin. The SAT and MRT are validated SpA measures whereas CT and Manikin are not SpA measures, *per se*. Further, the strength of relation between the BRT and the other validated tasks showed that the BRT is *at least* a very strong SpA measure.

Hypothesis II

General Discussion

In hypothesis II it was presumed that validated tasks would predict BRT performance. Models were provided in which it was shown that the SAT spatial orientation task was the single best predictor of BRT performance. With Manikin included, an additional 10% of BRT performance variance was explained; and with MRT performance included in a model containing SAT, an additional 4% of BRT performance variance was explained. This result is interesting given that Manikin was considered to be primarily a test of mental processing speed. To investigate the relationship between the time function and BRT performance, an hierarchical regression was performed using a variable consisting of mean BRT completion time. A model that included BRT item completion time only added 4% explanation in the variance of BRT performance, indicating that BRT item completion time was not necessarily a good predictor in relation to other tasks that were used in this study. However, mean BRT completion time was significantly correlated with BRT performance (*R* = .22), indicating that participants who took longer to complete an item generally outperformed those who do not.

Gender

The notion that females use different processing style to solve spatial problems is supported in this study given that males' and females' performance on the BRT could not be predicted using the same performance metrics. That is, female performance on the BRT was predicted by a variety of factors given that the predictive model of female performance included performance on all assessments that were used in the study.

Conversely, a single best predictor was shown in the male data, where the SAT, a measure of spatial orientation, proved to be a better predictor than all the assessments combined. However, as indicated by Model #2 for males, all assessments were shown to be a significant predictor of BRT performance as well.

These results may reflect that different processing styles can impact assessment of spatial ability as it is currently being conducted in the U.S. Navy. The results of the different predictive models produced by males and females with regard to BRT performance show that, at least in the instance of BRT performance, different approaches are used by the genders to solve spatial problems. Based on these results, the findings suggest that performance on the BRT can be explained by both simple and complex models whereas female performance can only be explained by a complex model. This finding supports the work of Bowers, Milham and Price (1998), who indicate that female spatial processing is more "distributed" than that of males.

Hypothesis III

The correlation between time and performance in the BRT model that has been given was significant. It shows that, on average, for every 0.47 msec there was a disparity reduction of 0.000010. This would be useful to know for any attempt at further validation, but based on data screening, findings suggest that the true chronometric relationship is more likely curvilinear, with an initial perturbation that often causes more error than error reduction, and a subsequent, generally "exponential" error reduction. The initial perturbation was presumably a means for participants to get their bearings, and to derive a better understanding of how the block arrays are configured.

After obtaining this information, participants were presumably better equipped to reduce error between the block arrays, as suggested by the rapid tapering off of disparity.

Data in the BRT chronometric model for the purposes of this study were constrained to a linear analysis. It is also possible, and likely, that the interaction between psychomotor and SpA processing is a non-linear function, exemplified, for example, by performance such as that depicted in Figure 7. However, a simple linear model is appropriate for the purpose of providing an initial glimpse of the BRT's psychometric properties.

CT data show that participants were generally adept at keeping the cursor in the general vicinity of the target, although the control reversal and variable movements assist in creating variability in the data.

Time and Performance

Notwithstanding the significant contribution of the "speed of processing" construct to performance in measures of general cognitive ability (Sternberg, 200), it is necessary to describe the relationship between time, spatial processing and psychomotor performance and how they relate to the results of this study. Having roots in Gibsonian ecological theory (1958) is Lee's (2006) theory of perceptuomotor control that includes the "ecological invariant" referred to as tau (τ). τ theory generally posits that the organism acts a unitary entity in dynamic relations with the environment, highlighting interaction between the organism and its environment, and does not consider reduction of the organism into analyzable parts.

τ theory development has been driven by a focus on ecological invariants in organism-environment interaction and is applied within the context of prospective guidance of movement, including "internal movements" (Lee, 2006). Internal movements are indicated by activation patterns in the sensory and in the nervous system.

A central tenet of the theory is that movement is guided by "τ -coupling motion-gaps," where conscious body movements, and associated cognitive processes, require guided closure of motion-gaps. Motion-gaps are defined as " the changing gap between a current state and a goal state," not unlike the condition imposed by the BRT of the current study. It is important to note for possible future reference that quantification of τ of a motion-gap is defined as the current size of the motion-gap divided by closure rate (Lee, 2006).

This approach to understanding the psychological phenomena that occur during task completion, such as those imposed by the PBMs of the current study, concords with a neuropsychological approach to understanding. Psychomotor movements being controlled by feedback control systems, for example, are no doubt influenced by a τ-type process, if not τ specifically. Sensory organs associated with feedback control systems permit comparison, at "higher levels," of stimuli to a reference by a "comparator," including the processing of τ-type data (i.e., an estimate of time-to contact) that could be used to modulate controlling mechanisms for error correction. Merchant and Georgopoulos (2006) provide a review and detailed descriptions of how τ plays a role in the neuropsychology of spatial and psychomotor processing.

In relation to the current experimental context, τ or a similar measure may prove to be a useful means by which novel PBM performance data are transformed in future studies to compare BRT performance data with that of a similar ecological performance domain (e.g., flight training).

Study Limitations

It should be noted that participants in this study were university students. Although all Navy flight students are college graduates, they constitute a qualitatively unique subpopulation in comparison to university students. It is not known whether the findings here would mirror the results of the same study in which U.S. Navy Aviation students participated.

Given that gender differences have been expressed in the data here, it is possible that, without corrective or compensatory action (e.g., weighting), that the BRT could be perceived as discriminatory based on gender.

Another limitation of this study is that it will not provide a wide spectrum of evidence regarding application of other performance-based measures as good predictors of flight training performance. The focus of this research is in regards to SpA only and does not apply to other factors such as general psychomotor ability. Further, this study identifies a specific aspect of SpA in a constrained context. The tasks that participants were asked to complete are tasks that are constrained to specific domains, primarily mental rotation for the BRT.

Recommendations

- It is strongly recommended that an item analysis be conducted of the BRT. Such an analysis can produce results indicating what items are more difficult than others. It could also indicate differential chronometric models based on block array complexity or angular disparity
- Complete an incremental validity study to determine if and to what degree the BRT-CT combination adds explanatory power to flight training performance to that of the ASTB
- Produce "cleaner" performance models for the BRT and CT tasks by means of gathering reliability data and determine if performance models (e.g., ideal vs. less than ideal) can be differentially classified for the BRT
- Determine contribution of psychomotor ability to BRT and CT tasks via comparison by using "purer" psychomotor tasks (e.g., finger tapping)
- Determine the utility of these tasks outside of U.S. Naval Aviation selection (e.g., diving community), and/or outside of the personnel selection arena (e.g., brain injury diagnosis and therapy)

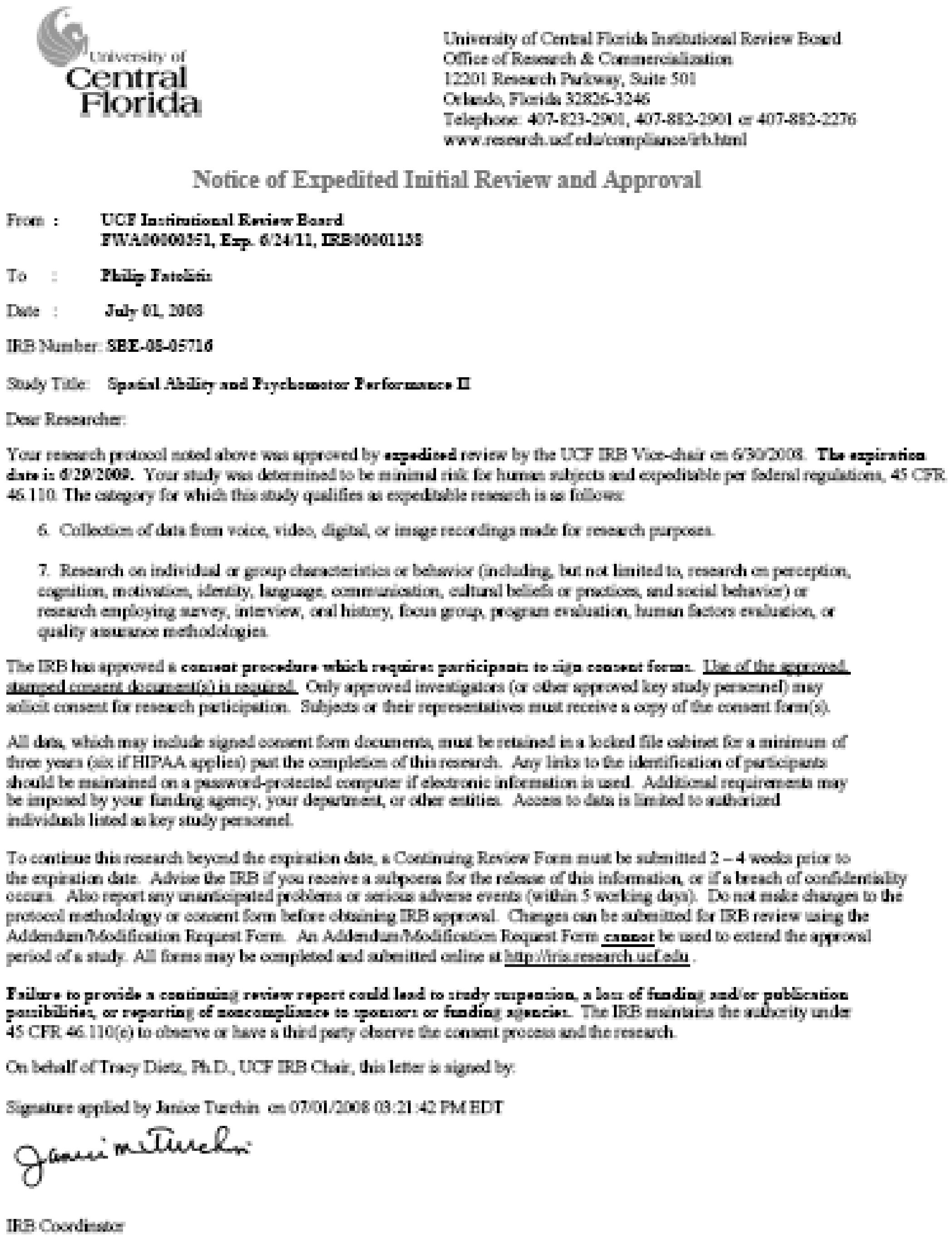
Conclusion

These findings could ultimately result in the utilization of a performance-based measure for selection U.S. Naval Aviation personnel for training. It was found that the experimental tasks, although not entirely in line with hypothetical reasoning, could constitute a good computer-based measure of spatial ability.

The research conducted here provides evidence in support of the notion that computerized, performance-based measures can measure constructs that are similar to those that are currently measured by paper-and-pencil tests. If support is eventually found in favor of increased ecological and predictive validity of the novel tasks introduced here, it will result in a significant, demonstrable savings in training dollars and possibly relate to increased Navy aviation safety.

A more general significant finding in this study is the performance algorithm of the BRT, similar to that of Shepard and Metzler (1971), where a strong linear relationship was found between response time and degrees of separation among mentally rotated stimuli. Shepard and Metzler also found that, on average, mental rotation takes place at the rate of 60° per second. Although not presented in terms of degrees, a similar human performance model was presented here and may be of some value to general knowledge.

APPENDIX A: IRB MATERIALS



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

Notice of Expedited Initial Review and Approval

From : **UCF Institutional Review Board**
FWA00000351, Exp. 6/24/11, IRB00001138

To : **Philip Fatolitis**

Date : **July 01, 2008**

IRB Number: **SBE-08-05716**

Study Title: **Spatial Ability and Psychomotor Performance II**

Dear Researcher:

Your research protocol noted above was approved by **expedited** review by the UCF IRB Vice-chair on 6/30/2008. **The expiration date is 6/29/2009.** Your study was determined to be minimal risk for human subjects and expeditable per federal regulations, 45 CFR 46.110. The category for which this study qualifies as expeditable research is as follows:

6. Collection of data from voice, video, digital, or image recordings made for research purposes.

7. Research on individual or group characteristics or behavior (including, but not limited to, research on perception, cognition, motivation, identity, language, communication, cultural beliefs or practices, and social behavior) or research employing survey, interview, oral history, focus group, program evaluation, human factors evaluation, or quality assurance methodologies.

The IRB has approved a **consent procedure which requires participants to sign consent forms.** Use of the approved, stamped consent document(s) is required. Only approved investigators (or other approved key study personnel) may solicit consent for research participation. Subjects or their representatives must receive a copy of the consent form(s).

All data, which may include signed consent form documents, must be retained in a locked file cabinet for a minimum of three years (six if HIPAA applies) past the completion of this research. Any links to the identification of participants should be maintained on a password-protected computer if electronic information is used. Additional requirements may be imposed by your funding agency, your department, or other entities. Access to data is limited to authorized individuals listed as key study personnel.

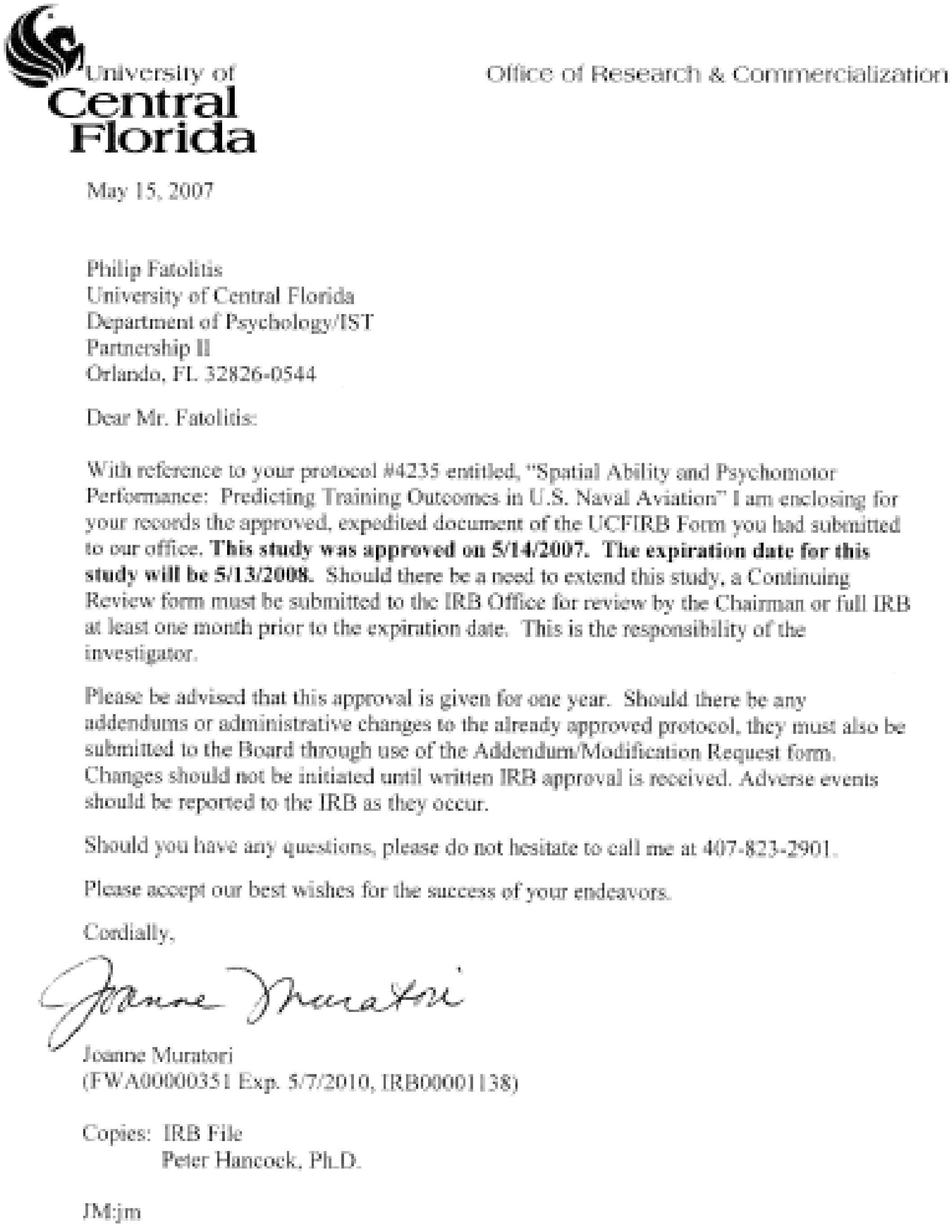
To continue this research beyond the expiration date, a Continuing Review Form must be submitted 2 – 4 weeks prior to the expiration date. Advise the IRB if you receive a subpoena for the release of this information, or if a breach of confidentiality occurs. Also report any unanticipated problems or serious adverse events (within 5 working days). Do not make changes to the protocol methodology or consent form before obtaining IRB approval. Changes can be submitted for IRB review using the Addendum/Modification Request Form. An Addendum/Modification Request Form **cannot** be used to extend the approval period of a study. All forms may be completed and submitted online at http://iris.research.ucf.edu .

Failure to provide a continuing review report could lead to study suspension, a loss of funding and/or publication possibilities, or reporting of noncompliance to sponsors or funding agencies. The IRB maintains the authority under 45 CFR 46.110(e) to observe or have a third party observe the consent process and the research.

On behalf of Tracy Dietz, Ph.D., UCF IRB Chair, this letter is signed by:

Signature applied by Janice Turchin on 07/01/2008 03:21:42 PM EDT

IRB Coordinator



University of Central Florida

Office of Research & Commercialization

May 15, 2007

Philip Fatolitis
University of Central Florida
Department of Psychology/IST
Partnership II
Orlando, FL 32826-0544

Dear Mr. Fatolitis:

With reference to your protocol #4235 entitled, "Spatial Ability and Psychomotor Performance: Predicting Training Outcomes in U.S. Naval Aviation" I am enclosing for your records the approved, expedited document of the UCFIRB Form you had submitted to our office. **This study was approved on 5/14/2007. The expiration date for this study will be 5/13/2008.** Should there be a need to extend this study, a Continuing Review form must be submitted to the IRB Office for review by the Chairman or full IRB at least one month prior to the expiration date. This is the responsibility of the investigator.

Please be advised that this approval is given for one year. Should there be any addendums or administrative changes to the already approved protocol, they must also be submitted to the Board through use of the Addendum/Modification Request form. Changes should not be initiated until written IRB approval is received. Adverse events should be reported to the IRB as they occur.

Should you have any questions, please do not hesitate to call me at 407-823-2901.

Please accept our best wishes for the success of your endeavors.

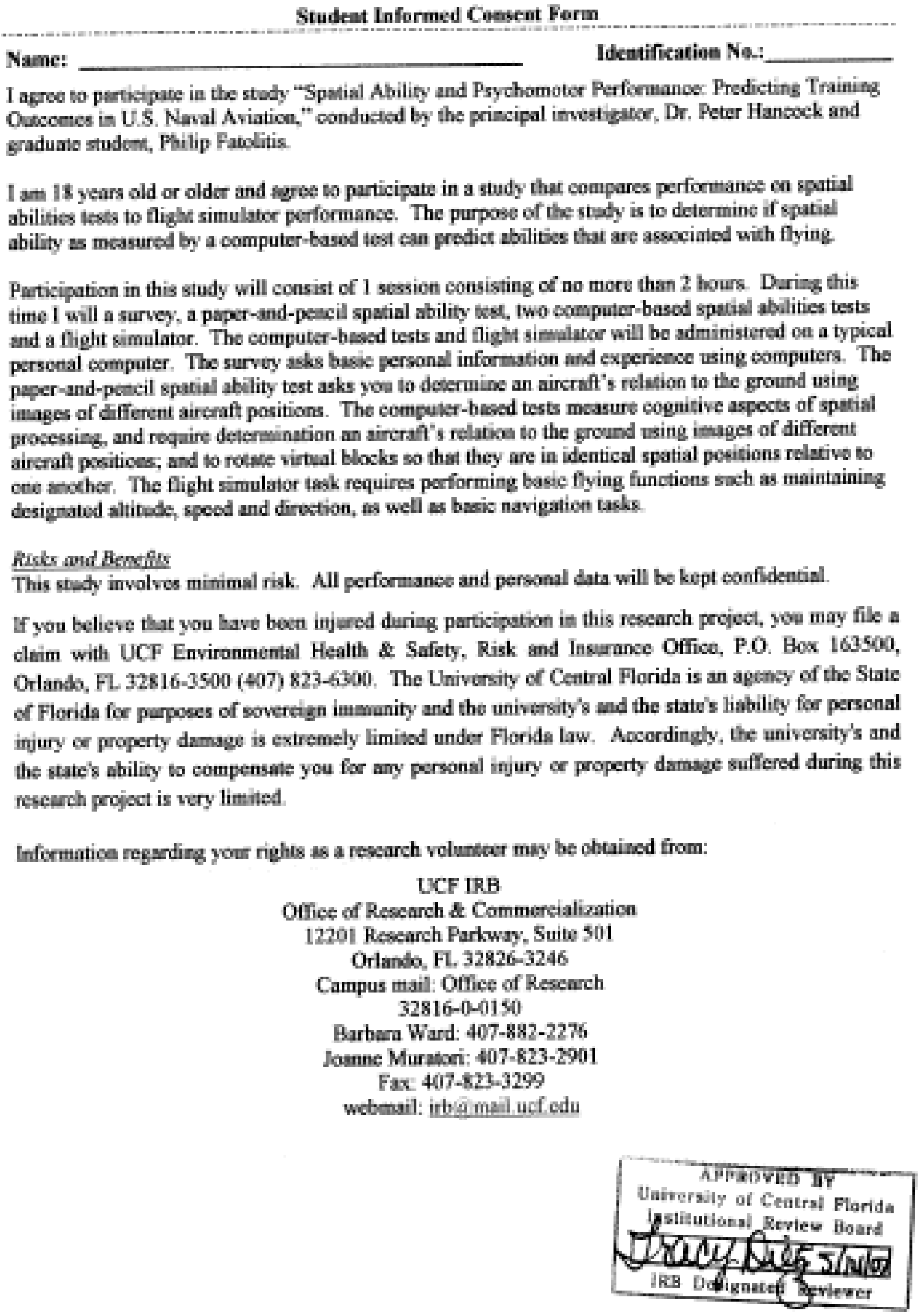
Cordially,

Joanne Muratori

Joanne Muratori
(FWA00000351 Exp. 5/7/2010, IRB00001138)

Copies: IRB File
Peter Hancock, Ph.D.

JM:jm



Student Informed Consent Form

Name: ______________________________ **Identification No.:**___________

I agree to participate in the study "Spatial Ability and Psychomotor Performance: Predicting Training Outcomes in U.S. Naval Aviation," conducted by the principal investigator, Dr. Peter Hancock and graduate student, Philip Fatolitis.

I am 18 years old or older and agree to participate in a study that compares performance on spatial abilities tests to flight simulator performance. The purpose of the study is to determine if spatial ability as measured by a computer-based test can predict abilities that are associated with flying.

Participation in this study will consist of 1 session consisting of no more than 2 hours. During this time I will a survey, a paper-and-pencil spatial ability test, two computer-based spatial abilities tests and a flight simulator. The computer-based tests and flight simulator will be administered on a typical personal computer. The survey asks basic personal information and experience using computers. The paper-and-pencil spatial ability test asks you to determine an aircraft's relation to the ground using images of different aircraft positions. The computer-based tests measure cognitive aspects of spatial processing, and require determination an aircraft's relation to the ground using images of different aircraft positions; and to rotate virtual blocks so that they are in identical spatial positions relative to one another. The flight simulator task requires performing basic flying functions such as maintaining designated altitude, speed and direction, as well as basic navigation tasks.

Risks and Benefits
This study involves minimal risk. All performance and personal data will be kept confidential.

If you believe that you have been injured during participation in this research project, you may file a claim with UCF Environmental Health & Safety, Risk and Insurance Office, P.O. Box 163500, Orlando, FL 32816-3500 (407) 823-6300. The University of Central Florida is an agency of the State of Florida for purposes of sovereign immunity and the university's and the state's liability for personal injury or property damage is extremely limited under Florida law. Accordingly, the university's and the state's ability to compensate you for any personal injury or property damage suffered during this research project is very limited.

Information regarding your rights as a research volunteer may be obtained from:

UCF IRB
Office of Research & Commercialization
12201 Research Parkway, Suite 501
Orlando, FL 32826-3246
Campus mail: Office of Research
32816-0-0150
Barbara Ward: 407-882-2276
Joanne Muratori: 407-823-2901
Fax: 407-823-3299
webmail: irb@mail.ucf.edu

APPROVED BY
University of Central Florida
Institutional Review Board
5/16/07
IRB Designated Reviewer

Confidentiality of Personal Data:

All data I will contribute to this study will be held in strict confidentiality by the researchers. That is, my individual data will not be revealed to anyone other than the researchers and their immediate assistants.

To insure confidentiality, the following steps will be taken: (a) only researchers will have access to the data in paper or electronic form: (b) data will be stored in locked facilities; (c) the data being used will be identified by number instead of name; (c) information regarding name and personal information will be stored separately from data for analysis.

MY PARTICIPATION IN THIS RESEARCH IS VOLUNTARY AND I WILL NOT BE COMPENSATED. I CAN WITHDRAW MY PARTICIPATION AT ANY TIME WITHOUT PENALTY - THIS INCLUDES REMOVAL/DELETION OF ANY DATA I MAY HAVE CONTRIBUTED.

This research is conducted by Peter Hancock and Philip Fatolitis. I have been given the opportunity to ask the research assistants any questions I may have. For further questions regarding this research, contact Dr. Hancock or Philip Fatolitis:

Dr. Peter Hancock
Partnership II
Suite 337
Orlando, FL 32826-0544
Phone: (407) 823-2310
e-mail: phancock@pegasus.cc.ucf.edu

Philip Fatolitis
Partnership II
Suite 337
Orlando, FL 32826-0544
Phone: (407) 384-7384
e-mail: philip.fatolitis@navy.mil
lyn.phil@yahoo.com

Signature:______________________________ Date: ________________

Initial These Statements:

"If I begin to feel something like sea-sickness, motion sickness or simulator sickness I shall discontinue participating immediately and inform the experimenter" __________

"I understand that I can withdraw from this study at any time" __________

"A choice to withdraw from the study will not influence any grades or personal treatment" __________

APPROVED BY
University of Central Florida
Institutional Review Board
5/14/07
IRB Designated Reviewer

APPENDIX B: PRE-SIMULATION SURVEY

Pre-Simulation Survey

Participant Number:

Age:

Gender:

1. Have you ever piloted an aircraft before? Yes/No

2. If you answered *No* to question 1, skip to question 3.

   a) Do you have a pilot's license? Yes/No

   b) How long has it been since you've gotten your license (years)?

   c) What types of aircraft have you flown?

   d) How many hours have you logged?

   e) When was the last time that you've flown (please give approximate month/day/year)?

   f) How would you describe your skill as a pilot (1 = poor, 7 = excellent):

1 2 3 4 5 6 7

3. How many hours of gaming experience do you have?

4. How often do you play video games (everyday, once a week/month/year)?

5. When was the last time that you played a video game?

6. How would you describe your skill in using video games (1 = poor, 7 = excellent)?

1 2 3 4 5 6 7

7. How many hours of sleep did you get last night?

8. How would you describe your level of alertness today?

1 2 3 4 5 6 7

9. Do you have any comments relating to the experiment (general conditions, experiences, concerns, etc.)?

10. What is your dominant hand (left or right)?

REFERENCES

Ackerman, P.L. (2001). New approaches to assessment and evaluation of perceptual speed abilities. *Georgia Institute of Technology School of Psychology Technical Report DTIC ADA419862:* Atlanta.

Ackerman, P.L. & 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.L. & Lohman, D.F. (1990). An investigation of the effect of practice on the validity of spatial tests. *Final Report (NPRDC Contract N66001-88C-0291).* Minneapolis, MN: Personnel Decisions Research Institute.

ASTB. (2006). Retrieved March 3, 2006 from http://www.usnavy.vt.edu/CurrentStudents/Forms/astboverview.pdf

Binet, A. & Simon, T. (1916). *The Development of Intelligence in Children* (E.S. Kite, Trans.). Baltimore: Williams and Wilkens.

Buchner, A., Erdfelder, E. & Faul, F. (2001). G-Power 2.0. Retrieved from http://www.psycho.uni-duesseldorf.de/abteilungen/aap/gpower3/ on 6 June 2008.

Carretta, T.R. (1989). USAF pilot selection and classification systems. *Aviation, Space, and Environmental Medicine, 60,* 46-49.

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

Cohen, J. (1992). A power primer. *Psychological Bulletin, 112,* 155-159.

Delaney, H.D. (1990). Validation of dichotic listening and psychomotor task performance as predictors of primary flight training criteria: Highlighting relevant statistical issues. *Technical Report No. AD-A227 010*, Naval Aerospace Medical Research Laboratory.

Fatolitis, P.G. (2007). *Spatial ability and psychomotor performance: Preliminary report on the validation of novel performance-based measures.* Unpublished manuscript.

Faul, F. & Erdfelder, E. (1992). GPOWER: A priori, post-hoc and compromise power analysis for MS-DOS (computer program). Bonn, FRG: Bonn University, Department of Psychology.

Fleishman, E.A. (1956). Psychomotor selection tests: Research and application in the united states air force. *Personel Psychology, 9,* 449-467.

Ghez, C. & Krakauer, J. (2000). The organization of movement. In E.R. Kandel, J.H. Schwartz & T.M. Jessell (Eds.), *Principles of Neural Science* (4th ed.) (pp. 653-673). New York: McGraw-Hill.

Gibson, J.J. (1958). Visually controlled locomotion and visual orientation in animals. *British Journal of Psychology, 49,* 182-194.

Griffin, G.R. (1987). Development and evaluation of an automated series of single-and multiple-dichotic listening and psychomotor tasks. *Technical Report No. AD-A199490*, Naval Aerospace Medical Research Laboratory.

Guilford, J.P. (1958). A system of the psychomotor abilities. *American Journal of Psychology, 71(1),* 164-174.

Guilford, J.P. & Lacey, J.I. (Eds) (1947). Printed classification tests. *AAF Aviation Psychology Research Program Reports (No. 5).* Washinton D.C.: GPO.

Guilford, J. P., & Zimmerman, W. S. (1948). The Guilford-Zimmerman aptitude survey. *Journal of Applied Psychology, 32(1),* 24-35.

Johnston, P.J. & Catano, V.M. (2002). Psychomotor abilities tests as predictors of training performance. Canadian Journal of Behavioural Science, 34, 75-83.

Kandel, E.R., Kupferman, I., & Iversen, S. (2000). Learning and Memory. In E.R. Kandel, J.H. Schwartz & T.M. Jessell (Eds.), *Principles of Neural Science* (4th ed.) (pp. 1227-1245). New York: McGraw-Hill.

Kandel, E.R. & Wurtz, R.H. (2000). Constructing the visual image. In E.R. Kandel, J.H. Schwartz & T.M. Jessell (Eds.), *Principles of Neural Science* (4th ed.) (pp. 492-506). New York: McGraw-Hill.

Lane, N.E. & Kennedy, R.S. (1990). Users manual for the automated performance test system (APTS). *NASA-CR-185631*. Houston, TX: NASA Johnson Space Center.

Lee, D.N. (2006). How movement is guided. Retrieved August 20, 2008 from http://www.perception-in-action.ed.ac.uk/publications.htm

Levin, S.L., Mohamed, F.B., & Platek, S.M. (2005). Common ground for spatial cognition? A behavioral and fMRI study of sex differences in mental rotation and spatial working memory. *Evolutionary Psychology, 3,* 227-254.

Lohman, D.F. (2000). Complex information processing and intelligence. In R.J. Sternberg (Ed.), *Handbook of Intelligence*. Cambridge: Cambridge University Press.

Lohman, D.F. (1993, July). *Spatial Ability and G.* Paper presented at the first Spearman Seminar, University of Plymouth, UK.

Lohman, D. (1979). Spatial ability: A review and reanalysis of the correlational literature. *Technical Report, No. 8.,* Stanford University, Aptitude Research Project, School of Education.

Lohman, D. F., Pellegrino, J. W., Alderton, D. L., & Regian, J. W. (1987). Dimensions and components of individual differences in spatial abilities. In S. H. Irvine & S. E. Newstead (Eds.), *Intelligence and Cognition: Contemporary Frames of Reference* (pp. 253-312). Dordrecht, The Netherlands: Martinus Nijhoff.

Lubinski, D. (2004). Introduction to the special section on cognitive abilities: 100 years after Spearman's (1904) "'General Intelligence,' Objectively Determined and Measured." *Journal of Personality and Social Psychology, 86,* 96-111.

Merchant, H. & Georgopoulos, A.P. (2006). Neurophysiology of perceptual and motor aspects of interception. *Journal of Neurophysiology, 95,* 1-13.

Naval Operational Medicine Institute. (2004). APEX.NET examiner's guide (NOMI P5098-C Part II). Pensacola, FL: Naval Operational Medicine Institute.

Pellegrino, J., Hunt, E., Abate, R., & Farr, S. (1987). A computer-based test battery for the assessment of static and dynamic spatial reasoning abilities. *Behavior Research Methods, Instruments, and Computers, 19,* 231-236.

Qu, S. (2003). Development and Testing of an Advanced Terrain Awareness and Warning System. Accessed 10 Dec 2006 from: http://ocw.mit.edu/NR/rdonlyres/Aeronautics-and-Astronautics/16-622Fall2003/B8DD3827-99D6-4C06-94B7-73401A5CD1AF/0/qu_shen.pdf

Saper, C.B., Iversen, S. & Frackowiak, R. (2000). Integration of sensory and motor function: The association areas of the cerebral cortex and the cognitive capabilities of the brain. In E.R. Kandel, J.H. Schwartz & T.M. Jessell (Eds.), *Principles of Neural Science* (4th ed.) (pp. 1227-1245). New York: McGraw-Hill.

Shea, D.L., Lubinski, D., & Benbow, C.P. (2001). Importance of assessing spatial ability in intellectually talented young adolescents: A 20-year longitudinal study. *Journal of Educational Psychology, 93,* 604-614.

Shepard, S. & Metzler, D. (1988). Mental rotation: Effects of dimensionality of objects and type of task. *Journal of Experimental Psychology: Human Perception and Performance, 14(1),* 3-11.

Snow, R.E. (1996). Aptitude development and education. *Psychology, Public Policy, and Law, 3/4,* 536-560.

Sternberg, R.J. (2000). *Handbook of Intelligence*. Cambridge, U.K.: Cambridge University Press.

Thurstone, L.L. (1938). *Primary Mental Abilities*. Chicago: University of Chicago Press.

U.S. Navy Flight Surgeon's Manual. (1989). Retrieved April 2007 from: http://www.iiimef.usmc.mil/medical/FMF/FMFE/FMFEref/fs_man/CHAPTER%2012.html

Vandenberg, S.G. & Kuse, A.R. (1978). Mental rotations, a group test of three-dimensional spatial visualiztion. *Perceptual and Motor Skills, 47,* 599-604.