2002

Measurement design : developing reliable alternate forms of cognitive tests

Jolene M. Bischoff
San Jose State University

MEASUREMENT DESIGN:

DEVELOPING RELIABLE ALTERNATE FORMS OF COGNITIVE TESTS

A Thesis

Presented to

The Faculty of the Department of Psychology

San Jose State University

In Partial Fulfillment

Of the Requirements for the Degree

Master of Arts

by

Jolene M. Bischoff

December 2002

UMI Number: 1411602

ProQuest Information and Learning Company
300 North Zeeb Road
P.O. Box 1346
Ann Arbor, MI 48106-1346

APPROVED FOR THE DEPARTMENT OF PSYCHOLOGY

Kevin Jordan, Ph.D.
Professor of Psychology, San Jose State University
SJSU Project Director

Barbara Burian, Ph.D.
San Jose State University Foundation
Senior Research Associate, NASA Ames Research Center

Mark Van Selst, Ph.D.
Professor of Psychology, San Jose State University

APPROVED FOR THE UNIVERSITY

ABSTRACT

MEASUREMENT DESIGN:

DEVELOPING RELIABLE ALTERNATE FORMS OF COGNITIVE TESTS

by Jolene M. Bischoff

Abstract

This alternate forms study examined correlations of forty undergraduate university students' scores on 4 alternate versions of the Trail Making B test, and 3 alternate versions of 2 subtests (Digit Symbol-Coding and Letter-Number Sequencing) from the Wechsler Adult Intelligence Scale-III Edition. Reliability analyses confirmed moderate to strong significant correlations among the original and alternate versions of the Digit Symbol-Coding (DSC) and Letter-Number Sequencing subtests (LNS). Significant moderate and non-significant weak correlations were found among the original and alternate versions of the Trail Making B tests. The sufficiently reliable alternate forms of these cognitive tests allow for implementation in future flight operational studies involving the cognitive performance of pilots in relation to their nutrition and blood glucose levels.

ACKNOWLEDGMENTS

This was one opportunity that I would have never imagined for myself. To be able to learn so much from and participate in various research projects, as well as to have my work supported at NASA Ames Research Center, made possible by the Cooperative Agreement with the San Jose State University Foundation Grant 21-1614-2868, has continued to amaze me. Special appreciation goes to Kevin Jordan, who offered me this wonderful professional and personal opportunity, creating experiences that will remain with me for many years to come. None of this would have been possible without your efforts – you truly make a difference with the work that you do and the manner in which you do it. Thank you for your continued support, professionalism, and personal dedication to students. Additionally, I would like to thank those who I also have the highest respect and admiration for as mentors and guides, for their technical and, at times, not-so-technical (but much needed) support. I am grateful to Immanuel Barshi, Barbara Burian, and Mark Van Selst for supporting this project. Thank you Barb for your ideas and suggestions, kind words of support, and guidance along the way. Thank you Yuri Tada, Robert Mauro, Wayne Stone, Amy Wang, Esther Johnson, and Beth Little for sharing your time, patience, and energy with me. Finally, I would like to thank my parents, family, and friends for their constant love and support, for being there for me, and for sharing this with me. Thank you Christopher for always believing in me.

This experience has allowed me to see a project through from beginning to end, with much anticipation, grief, and joy, in the midst of it all. Yet, in the end, reaching an understanding that it is just this process of dedicated individual and collective work and effort that culminates in such a unique and personal sense of accomplishment.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Measurement Design:

Developing Reliable Alternate Forms of Cognitive Tests

Jolene M. Bischoff

San Jose State University

Running Head: RELIABILITY OF ALTERNATE FORMS

Footnotes

Requests for reprints should be sent to Jolene M. Bischoff, Department of Psychology, San Jose State University, San Jose, California 95192.

Abstract

This alternate forms study examined correlations of forty undergraduate university students' scores on 4 alternate versions of the Trail Making B test, and 3 alternate versions of 2 subtests (Digit Symbol-Coding and Letter-Number Sequencing) from the Wechsler Adult Intelligence Scale-III Edition. Reliability analyses confirmed moderate to strong significant correlations among the original and alternate versions of the Digit Symbol-Coding (DSC) and Letter-Number Sequencing subtests (LNS). Significant moderate and non-significant weak correlations were found among the original and alternate versions of the Trail Making B tests. The sufficiently reliable alternate forms of these cognitive tests allow for implementation in future flight operational studies involving the cognitive performance of pilots in relation to their nutrition and blood glucose levels.

Measurement Design: Developing Reliable Alternate Forms of Cognitive Tests

Psychological tests and measures have allowed researchers to investigate the performance of individuals in compromised cognitive and physiological states. Assessing the specific nature of various injuries to and impairments of the brain is made possible through use of these clinical and neuropsychological tests. Performance of individuals on these tests allows researchers to determine what, if any, effects may occur as a result of natural or experimental manipulations.

General brain functioning and performance on various tasks such as visual-motor dexterity, attention, processing speed, concentration, and ability to shift mental sets have been negatively affected during controlled blood sugar experiments. For instance, reduced blood sugar levels, specifically, hypoglycemia or "an abnormally low plasma glucose level that leads to symptoms of sympathetic nervous system stimulation or of central nervous system dysfunction" (Merck Manual, 2001) has been found to significantly impair cognitive functioning. Hypoglycemia's effects on the central nervous system include symptoms such as confusion, deficiencies in coordination, headaches, blurred vision, anxiety, and dizziness (Field, 1989).

The literature has shown support for the notion that hypoglycemia significantly compromises the performance of various cognitive processes (Gonder-Frederick, Cox, Driesen, Ryan, & Clarke, 1994). This condition has resulted in longer response times and lower scores on cognitive tests during compromised blood sugar levels or hypoglycemic episodes (Cox, Gonder-Frederick, Schroeder, Cryer, & Clarke, 1993; Gold, Deary, MacLeod, & Frier, 1995; Gold, Deary, MacLeod, Thomson, & Frier, 1995; Gold, MacLeod, Deary, & Frier, 1995; McCrimmon, Deary, & Frier, 1997; McCrimmon, Deary, Huntly, MacLeod, & Frier, 1996; Taylor & Rachman, 1988). These findings are not surprising, given that glucose is the only nutrient that the brain can convert into energy (Doraz, 1987).

These tasks have reflected impairments in decision-making processes (Blackman, Towle, Lewis, Spire, & Polonsky, 1990) and even influence basic driving skills in simulation research (Cox, Gonder-Frederick, Kovatchev, Julian, & Clarke, 2000). Driving performance of participants with Type 1 diabetes was significantly impaired during some relatively mild hypoglycemia levels, with off-road driving (driving across the midline), driving fast, and applying brakes more on the open road occurring more frequently than at euglycemia (normal blood sugar level).

Cox and colleagues (2000) found that participants, in induced hypoglycemic states, failed to stop at stop signs and crashed significantly more at sudden stops during the last fifteen minutes of a thirty minute simulated driving task. Also, at some point during these hypoglycemic episodes all driving demands during the simulation, such as speed limit changes, detours at stop sign intersections, encroaching fixed objects, and negotiating oncoming and cross traffic were affected. It is surprising that 43% of the significantly impaired drivers never pulled over or treated their conditions (i.e., took "corrective action"), with an available glucose drink. Significant impairments in cognitive performance tasks have been found not only in insulin dependent diabetes mellitus (IDDM) adult participants, but also in healthy adults when their blood sugar levels were even moderately altered (Stevens et al., 1989).

Aviation Safety Reporting System

The negative effects of low blood sugar level on cognitive processes may be most apparent when job performance is demanding and of a highly sensitive nature. Consider the field of aviation; cognitive failures may result in situations that are dangerous, if not possibly catastrophic. Several reports submitted by pilots to the Aviation Safety Reporting System (ASRS) identified low blood sugar levels as having negatively affected their performance.

The ASRS was established through a 1976 Memorandum of Agreement between the Federal Aviation Administration (FAA) and National Aeronautics and Space Administration (NASA) to improve aviation safety. This voluntary and confidential reporting system allows pilots, flight attendants, air traffic controllers, ground personnel, and others to provide narrative accounts of "unsafe" incidents, and waives possible fines/penalties against reporters. These accounts are collected and analyzed; deficiencies are investigated and improvements can be made, contributing to the wealth of human factors research and to the National Aviation System. However, data from this source are anecdotal that are not certified as accurate, complete, or standardized by the FAA, National Transportation Safety Board (NTSB), or NASA. Given the associated interpretive difficulties, a conservative approach to an incident reporter's claim of causality, implications, and recommendations must be considered.

Over seventy ASRS reports (Aviation Safety Reporting System, 2001a) utilizing relevant search terms (see Appendix A) relate negative in-flight performance (i.e., physiological, mental, and/or emotional functioning) to either lack of food, low blood sugar, and/or hunger. Fifty-four ASRS reports were included in a taxonomy table and were grouped into categories. Sixteen reports were excluded from the data set because they did not contain the category title word, or other terms associated with that category, in the narrative portion in which the situation was described. Reports in the "Other Difficulties" category did not contain the accepted terminology; however, were retained for their value of similarity to other reports within the taxonomy table. This taxonomy included the following categories: "Physiological Difficulties" (e.g., hands-shaking, headache, cold/perspiration, blurred vision, dizziness, feeling sick, sleepy, tired, fatigued), "Attentional Difficulties" (e.g., distracted, pre-occupied, diverted, unaware, problems concentrating), "Difficulty Thinking" (e.g., mind-tired/fatigued, mis-reading, mis-calculating), "Impaired Judgment" (decision-making), "Memory Difficulties" (e.g., short & long term, forgetting), "Decreased Vigilance" (complacency), "Lack of Coordination", "Emotional Difficulties", and "Other Difficulties" (un-categorized reports) (see Table 1).

Many reports fall into more than one of these categories; however, the majority of identified reports found are applicable to the categories of physiological difficulties, followed by attentional difficulties, difficulty thinking and, impaired judgment. Cox and his colleagues (Cox, Gonder-Frederick, Antoun, Cryer, & Clarke, 1993) studied recognition of hypoglycemic symptoms in diabetic patients. Interestingly, they found that the most frequently reported symptoms in rank order appeared as the following: "difficulty concentrating, trembling, uncoordinated, pounding heart, slowed thinking,

Table 1

Taxonomy Table of ASRS Reports

Physiological Difficulties	Attentional Difficulties	Other Difficulties	Difficulty Thinking	Impaired Judgment
101351	123033*	80930	110833*	180894*
108886	130747*	168249	126061*	215225*
110833*	132717*	199170	151393*	245988*
126061*	141750*	199721	187126*	360530*
130747*	180894*	309500	245988*	397036*
132717*	215225*	314510	302930*	
141750*	223650	327924	425650*	
151393*	249070*	367839		
151596*	262034	391270		
158062	275060*			
168334	279425*			
176976*	296526			
180894*	309724*			
187126*	316820			
201403	329500			
244891	336129*			
270653	358020*			
275060*	358698			
279425*	425308*			
297539*	425650*			
310970				
302930*				
309724*				
331800				
336129*				
344040*				
358020*				
360530*				
395826				
397036*				

Memory Difficulties	Decreased Vigilance	Lack of Coordination	Emotional Difficulties
123033*	151393*	249070*	302930*
151596*	245988*	279425*	410667*
176976*	358020*	425650*	
218068*	425308*		
242236			

Note. An asterisk (*) indicates duplicate reports within these nine categories. 54 total reports.

nervous/tense, and sweaty" (Cox et al., 1993, p. 526), somewhat resembling reports of physiological and attentional difficulties within the ASRS system.

Several instances of pilots associating cognitive performance impairments with low blood sugar levels or lack of "proper nutrition or sustenance" can be seen within general and commercial aviation. For example, a small aircraft in New Jersey went off the end of a runway and collided with a dirt pile, damaging the plane's wing tips and propeller. The general aviation pilot attributed this incident to low blood sugar. The pilot's "brain ceased to function normally...judgement clouded...cause...low blood sugar and oxygen deficiency to brain" (Report No. 278700, ASRS, 2001a). A flight in Michigan was involved with what is labeled as an "anomaly", or an unsafe or illegal event. The pilot flew in the wrong direction and airspace. The pilot noted in the report that, "...fatigue, stress, perhaps low blood sugar from not eating, all contributed to the mistakes I made. All seem to be a part of this job. I know what I did wrong, but what will I do next time?" (Report No. 294222, ASRS, 2001a). The pilot described, "the faulty thought process that led to the decision...the mental picture of where I was and where I was going was skewed with images of other areas". More serious consequences could involve incidents where flight crew (pilots and flight attendants) and possibly hundreds of passengers are involved.

Commercial pilots have also noted the detrimental effects of compromised blood sugar levels. One first officer commented, "I think low blood sugar played a big part. The Captain had not eaten in almost 12 hours. I was just plain out of it, and the symptoms were insidious" (Report No. 310970, ASRS, 2001a). The flight crew had overshot an assigned altitude during descent. Evans, Pernet, Lomas, Jones, and Amiel (2000) have addressed exactly this "insidious" nature of lowered blood sugar levels and its influence on mental capabilities. They investigated the speed of onset of the symptoms of acute hypoglycemia and concluded that these symptoms may preclude awareness. Additionally, Evans and his colleagues found that, "symptom generation, including the subjective feeling of abnormality, was delayed for up to 20 minutes after the onset of the hypoglycemia nadir and after the onset of detectable cognitive dysfunction" (2000, p. 896).

Miscommunication, due to low blood sugar, may also play a role in flight incidents. One commercial pilot failed to correctly brake at the gate, assuming a "chocks in" signal by the marshaller. Damage to the plane's radome underside occurred and the pilot reported that, "Contributing to my confusion was low blood sugar" (Report No. 371631, ASRS, 2001a). During another commercial flight, several clearances and headings were described as "confusing" and were deviated from, with the pilot specifically stressing that, "I am certain that low blood sugar was a factor in the crew's performance" (Report No. 367839, ASRS, 2001a).

Billings and Reynard (1984) found that human-error comprised more than eighty percent of all ASRS reports between 1976–1982. Thus, addressing issues that affect decision-making abilities, monitoring, comprehension, communication, performance and other related cognitive processes is highly relevant. Ford and colleagues (Ford, Jack, Crisp, & Sandusky, 1999) analyzed NTSB accidents from 1990-1996 and re-affirmed that human-error, specifically decision errors and skill based errors, frequently continued to be identified as components (or "causes") of accidents. The National Transportation Safety Board (NTSB) has identified and documented low blood sugar, lack of food, and/or hypoglycemia specifically, as contributing factors in three commercial airline accident reports (NTSB, 1976; NTSB, 1979; & NTSB, 1986) and four general aviation aircraft accident reports (NTSB, 1989; NTSB, 1992; NTSB, 1997; & NTSB, 1998). What should be noted is that, within aviation transportation accidents, an abnormal glucose level resulting in hypoglycemia "cannot currently be determined from postmortem specimens using existing procedures" (Canfield, Chaturvedi, Boren, Véronneau, & White, 2000, p. 1). This leads to the necessity of investigating and analyzing this potentially "insidious" incapacitation from a more pro-active rather than reactive approach.

Investigations into factors related to hypoglycemia during aviation accidents have been conducted, although not recently (Fisher & Atkinson, 1980; Gibbons, Plechus, Chandler, & Ellis, 1966; Harper & Kidera, 1973; Meyer, 1969; Powell, 1956; Powell, Carey, Brent, & Taylor, 1957; Raichle & King, 1972). These studies have attempted to address, diagnose, and/or link pilots' poor flight performance (including periodic episodes of unconsciousness) with hypoglycemia resulting from poor nutritional habits, and/or alcohol-induced hypoglycemia (AIH). Recommendations were made to educate pilots about their dietary habits and nutritional needs.

Over the years several technical reports and conference papers, and few current journal articles have specifically addressed or experimentally tested pilot's flight incapacitation and accident potential primarily related to low blood sugar levels. A study conducted by Farrace and colleagues (Farrace, Urbani, Sakara, & De Angelis, 1993) investigated reactive hypoglycemia (resulting from carbohydrate ingestion) in a healthy population of military Air Force trainees. Finding a relatively high incidence of reported hypoglycemic symptoms by the trainees, they too, had dietary recommendations for the operational environment. Porcu and colleagues (Porcu, Berti, & Lala, 1988) studied EEGs (electroencephalogram) in a small sample of normal subjects in order to determine what effect hypoglycemia had on the central nervous system (CNS). They concluded that neuroglycopenic symptoms (a lack of glucose to the neurons) and its negative effect on the CNS were evident, which could be implicated in flight accidents. Other studies have also investigated flight accidents and the impaired performance of pilots due to compromised blood sugar levels; however, researchers have approached this issue either through postmortem samples, or crash analysis (Canfield et al., 2000; Guohua, 1994).

Pilot Nutrition Research

A research project at NASA Ames Research Center is currently focusing on issues involving the effects of low blood sugar levels on commercial pilots' performance. Primarily, researchers are interested in flight operations such as those of commuter airlines where pilots have busy days, short flights, quick turn-arounds, and no food provided to them onboard the aircraft. This creates a situation where commercial pilots may not be consuming the food that they need, because they are scheduled to work before food services at restaurants or airport terminals are open, or after they are closed.

Some data have already been collected from pilots regarding what they are eating (i.e., meal, snack, beverage) and how they are feeling (i.e., sleepy, dizzy, hungry, headache, etc.) when they are flying. This preliminary research collected daily nutritional information from commercial pilots via daily food logs, and is currently being analyzed. These logs included information on pilots' on/off duty times, legs of flight, flying and commuting times, types of meals, snacks and beverages consumed and physical symptoms. Incidents and accidents that have identified food (or lack of) as a factor are also being investigated.

The Need for Reliable Alternate Forms

Before conducting the planned assessment of cognitive impairment resulting from low blood sugar in the aviation operational environment, reliable alternate forms of tests used in the research environment need to be created. Smaller subject pools are utilized within the aviation environment. Therefore, employing a within subjects repeated measure design allows more data to be collected, lending to its efficiency and economy. Establishing reliable alternate forms of these specific tests will enable researchers to determine if, how, and at what point pilots' cognitive performance is affected in relation to their nutrition and blood sugar levels in subsequent flight simulator studies. The proposed simulator study will investigate compromised blood sugar levels and assess cognitive performance of pilots. These alternate forms will allow researchers to measure certain cognitive processes multiple times without the constraints of practice or carryover effects, which have significant tendencies to emerge in normal participants, especially in timed tests (Lezak, 1995). Creation of reliable alternate forms of relevant measures, often used in low blood sugar research, will allow for a "truer" performance score and investigation into the possibility of pilots' operational impairment, and compromise of aviation safety.

The flight simulator study will collect cognitive performance data from commercial pilots in normal cockpit tasks and during various scenarios. Pilots will be observed and tested on two separate occasions. Pre- and post-test measures will be given across the two simulator trials to assess cognitive performance. Again, since practice and carryover effects were considered in using one test version repeatedly, reliable alternate test versions are necessary. We have decided that three alternative versions of two performance tests (Digit Symbol-Coding and Letter-Number Sequencing) and four alternative versions of an additional performance test (Trail Making B) will be needed. These tests each measure specific cognitive properties and will be easy to administer during the simulator trials, taking a total of about twenty minutes.

Test Development

Specific tests have been reviewed in experimental research to assess the cognitive performance deficits found in individuals (with and without diabetes), due to compromised blood sugar levels (Bischoff, Warzak, Maguire, & Corley, 1992). These tests have detected significant cognitive deficits in individuals and have been often used and adopted in blood glucose research. Descriptions of some of the tests utilized in blood glucose research and their assessment of various cognitive processes can be found in Table 2 (see Table 2).

One test commonly used to assess cognitive impairments due to hypoglycemia in normal subjects has been the Digit Symbol-Coding subtest from the Wechsler Adult

Table 2

Cognitive Processes & Readily Used Performance Tests

	WAIS-III	Trail Making B	PASAT	Cog-Screen
Non-verbal Comprehension	PC			
Contextual Inference	V			
Visual Motor Dexterity	DSC			
Persistence in the Face of a Boring Task	DSC			
Clerical Speed	DSC			
Anxiety Indicator	DSC, A			
Distractibility	DSC, A, DS, LNS		X	BDS, MP
Persistence	DSC			
Working Under Time Pressure	DSC, BD, A, PA, SS			
Verbal Concept Formation	S, OA			
Analysis of Spatial Relations	BD			
Rigid Application of Logic	BD			
Working Memory	BD, A, DS (auditory, immediate), DF, DB, LNS		X	BDS, MP (long-term), VSC, SDC, DA, P
Reasoning	BD, A (non-verbal, inductive)			MP (logical), SA (deductive)
Problem Solving	BD			
Perceptual Organization	BD, DB (mental reordering)		X	
Auditory Concentration	A, SP, LNS			
Concentration	A, LNS	X	X	BDS, MP
Attention Span	A, DS, LNS		X	BDS, MP, VSC, SDC, DA (divided), P

Pattern Completion	MR			
Classification	MR			
Alertness to Environment	I			
Foreign Language Background	I			
School Learning	DSC, A, I		X (arithmetic ability)	MP (arithmetic ability)
Cultural Opportunities	I, PC, C			
Sequential Thinking	PA			
Cognitive Processing Speed	DSC, SS	X	X	

Readily used performance tests (Trail Making B & PASAT) and subtests of tests (WAIS-III & Abbreviated version of the Cog Screen). Adapted from "Foundations of Intellectual Assessment: The WAIS-III and Other Tests in Clinical Practice", by R.J. Gregory, Copyright 1999 by Allyn and Bacon.

Note. (WAIS-III sub tests): Picture Completion = PC, Vocabulary = V, Digit Symbol-Coding = DSC, Similarities = S, Block Design = BD, Arithmetic = A, Matrix Reasoning = MR, Digit Span = DS, Digit Span Forward = DSF, Digit Span Backward = DSB, Information = I, Picture Arrangement = PA, Comprehension = C, Symbol Search = SS, Letter-Number Sequencing = LNS, Object Assembly = OA

(CogScreen sub tests): Backward Digit Span = BDS, Math Problems = MP, Visual Sequence Comparison = VSC, Symbol Digit Coding = SDC, Divided Attention = DA, Pathfinder = P, Shifting Attention = SA

Intelligence Scale–Third Edition (WAIS-III; Gold, Deary, MacLeod, & Frier, 1995; Gold, Deary, MacLeod, Thomson et al., 1995; McCrimmon, et al., 1997; Mc Crimmon et al., 1996). The Digit Symbol-Coding subtest (DSC), also known as the Digit Symbol Substitution Task (DSST) is comprised of nine digits, or numbers that are represented by nine different symbols. The participant is required to copy the correct symbol (using a coding key) for a given number (ranging from 1 to 9). The participant draws the appropriate symbol in a box under its corresponding number. There are 133 boxes to fill in with symbols and the participant is given a 120-second time limit. The score is determined from the number of symbols drawn correctly within that time limit.

Alternate versions of the Digit Symbol-Coding subtest, and the Trail Making B test have been developed. The Trail Making B test, requires participants to correctly connect twenty-five circles enclosed with numbers from 1 to 13, and letters from A to L that are scattered on an 8 1/2 x 11" page, in an alternating sequence (i.e., 1-A, 2-B, 3-C, etc…), as quickly as possible. Lewis and Rennick (1979) created the Repeatable Cognitive–Perceptual–Motor Battery (RCPM), with subtests that include alternate forms of the Trail Making B, Digit Symbol, Digit Vigilance, and Visual Search tests. This testing battery exhibited high test-retest correlations and alternate form reliability for the Digit Symbol and Trails B test, at .87 and .80 respectively (Kelland & Lewis, 1994). Overall, the RCPM has reached satisfactory reliability with the original formats (Lezak, 1995).

The Symbol Digit Modalities Test (SDMT) developed by Smith (1973) is another alternate version of the Digit Symbol-Coding subtest, which reverses the presentation of numbers and symbols (Lezak, 1995). While the DSC subtest requires participants to write in the symbols that go with each number, the SDMT requires participants to write-in numbers that go with each symbol. This written version of the SDMT correlates highly with the Digit Symbol subtest, and has been used in conjunction with an oral administration version of the SDMT.

The Trail Making B test and the Paced Auditory Serial Addition Task (PASAT) have also been frequently used as cognitive assessment tools in hypoglycemia studies. The PASAT requires a participant to listen to an audio tape of fixed random sequences of numbers 1– 9 and add the first number to the second number, give the answer, add the second number to the third and give the answer, and so on for sixty–one numbers. The score is the total number of correct answers on each run (maximum score of 60). Although the PASAT has been used in blood glucose studies, the length of time for test administration (minimum 20 minutes) and its high correlation to WAIS scores (Gold et al., 1995a) do not make it a useful choice for creating a brief alternate form for in this study.

Stanczak (1986) created a modified version of the Trail Making B test, called the Expanded Trail Making B test (EMMT), which uses clock faces (Part X), or increasing dots sizes (Part Y) for participants to connect in order as a timed test. The ETMT has retained significant correlations with forms A and B of the Trail Making B test evidencing concurrent validity (Stanczak, Lynch, McNeil, & Brown, 1998). However, there is limited empirical research on the reliability of the ETMT test (Lezak, 1995).

The Color Trails 2 created by Maj et al. (1993) was initially created as an unbiased culture-free version of the Trail Making B test, for participants unfamiliar with the English alphabet. This non-alphabetical parallel form has twenty-five scattered circles on a page, but has each color set (pink or yellow) numbered. Odd-numbered circles are yellow (numbered up to 13) and even-numbered circles are pink (numbered up to 12). Participants must alternate between yellow then, pink in ascending numerical order (1Y to 2P, to 3Y, etc.). This alternate version correlates with the TMB at .50 (Lezak, 1995).

The Letter-Number Sequencing subtest (of the WAIS-III) requires that a participant listen to a sequence of numbers and letters, read by the researcher at a one-second interval pace. Participants hear a combination of letters and numbers and are asked after each sequence to repeat the numbers first, in order, starting with the lowest number and then the letters in alphabetical order. There are seven trial items (each item consists of three trials) consisting of increasing number and letter combinations, from two items up to eight items. This subtest has not yet been utilized within the low blood glucose research. However, interest exists in the ability to use the test as an assessment of auditory concentration, working memory, and ability to shift mental sets. The cognitive tasks measured by this subtest are those that comprise the fundamental tasks performed in the operational aviation environment, such as interpretation and integration of a variety of sensory information, which makes this measure highly relevant for the current study.

General brain functioning and (visual and auditory) sensory processing speed have been assessed in relation to acute hypoglycemic states using the DSC test (Gold et al., 1995a, 1995b, 1995c; McCrimmon et al., 1997; Mc Crimmon et al., 1996;), the Trail Making B test (Gold et al., 1995a, 1995b, 1995c; Hoffman et al., 1989; McCrimmon et al., 1997; Mc Crimmon et al., 1996), the PASAT (Cox, Gonder-Frederick, Antoun, et al., 1993; Gold et al., 1995a, 1995c) or a combination of these. Significant impairments in subjects' scores (lower scores and longer response time) on these tasks due to induced hypoglycemia were found in all but two studies for the Trail Making B test (Gold et al., 1995a, 1995b). These studies have not been geared directly towards aviation-related cognitive impairments, but do assess similar cognitive processes such as visual motor integration and mental flexibility.

By isolating specific cognitive tasks, three alternate versions of the two subtests (Digit Symbol-Coding and Letter-Number Sequencing) and four alternative versions of the Trail Making B test were developed for this study, resulting in a basic time-efficient testing device. These alternate forms will assess relevant auditory and visual processes; both processes have been found to be negatively affected by blood sugar levels (Mc Crimmon et al., 1997; Mc Crimmon et al., 1996; Tabandeh, Ranganath, & Marks, 1996) and may play significant roles within flight operations.

Since piloting an aircraft is a complex task, it requires unimpaired sensory perception. Memorization, judgement, communication, coordination, and motor skill can be very demanding tasks during aircraft maneuvering, readback and hearback, exposure to visual illusions, and spatial orientation, requiring a keen sense of sight and sound.

Billings and Reynard (1984) state that, "...vision plays a large part in avoiding mid-air collisions and that the 'see-and-avoid' concept is an effective preventative" (p. 964). Mc Crimmon and colleagues (1996) investigated particular aspects of visual perception in non-diabetic participants and found that while insulin-induced hypoglycemia did not appear to affect visual acuity or stereoscopic vision, significant deterioration occurred in the ability to detect differences of contrast. Participants' cognitive performance was negatively affected, visual inspection time was longer, and ability to detect visual change and movement was impaired. Similarly, Mc Crimmon and colleagues (1997) found significant deterioration in non-diabetics "simple" auditory temporal processing and loudness discrimination. Hypoglycemia was also shown to significantly disrupt short-term auditory storage and slow the speed of information processing.

Considering these findings and the comparable tasks required on the flightdeck, visual processes will be assessed using newly-created alternate forms of the Digit Symbol-Coding subtest and of the modified version of the Trail Making B test. These tests both assess cognitive processing speed and visual motor integration. Independently, the Trail Making B test assesses visual concentration and the ability to shift mental sets while the Digit Symbol-Coding subtest assesses visual-motor dexterity, persistence in the face of a boring task, and clerical speed. This subtest can be influenced by participants' anxiety, distractibility, and ability to work under time pressure (Gregory, 1999). Auditory processes will be assessed using constructed alternate forms of the Letter-Number Sequencing subtest (of the WAIS-III). This test assesses attention span, distractibility, and working memory. Letter-Number Sequencing also assesses auditory concentration and the ability to shift mental sets.

Construction of Alternate Forms

In order to create reliable alternate forms of pre-existing tests, adherence to the tests' original layout, directions, and difficulty levels (except for the Trail Making B test) was attempted. The layout of Digit Symbol-Coding remained similar to the original, with the same number of rows, columns, and boxes on the testing sheet and with the coding key at the top. The majority of coding key numbers, 1 through 9, were represented with entirely new symbols (Liungman, 1991). Certain symbols were rotated in some cases. Symbols (not letters) were chosen for their similar number of pencil strokes (i.e., lifting, direction, and movement of the pencil) and simplicity. The same number of seven practice sample items remained. Original numbers from the DSC coding key were re-numbered for the alternate versions. This method was based upon referencing the original DSC numbers. That is, in re-numbering the boxes on version one, a 1 on the original version was changed to a 2 on the alternate version, a 2 was changed to a 3, a 3 was changed to a 4, 4 to7, 5 to 6, 6 to 5, 7 to 9, 8 to 1, and 9 to 8. For version two, a 1 was changed to a 4, 2 to 8, 3 to 9, 4 to 6, 5 to 2, 6 to 1, 7 to 3, 8 to 5, and 9 to 7. For version three, a 1 was changed to a 3, 2 to 6, 3 to 2, 4 to 1, 5 to 9, 6 to 8, 7 to 4, 8 to 7, and 9 to 5. This would assure that "repeaters" or numbers that appear repeatedly and more than other numbers during the first two to three rows were mimicked on these alternate forms.

The construction of the Letter-Number Sequencing test (LNS), which was administered verbally, was modified in terms of changing the numbers and letters. The original LNS test does not include the letters I, O, or U as part of the possible combinations. It does include repeated use of numbers 1 through 9. In creating alternate versions, these practices were also followed. A similar ratio of numbers to letters used for alternate versions was considered. The reading layout (for the test administrator) of the original LNS test sequence of reading the letter then number or the number then letter, in an alternating succession was also maintained.

The layout of the alternate versions of the Trail Making B and its practice tests remain relatively similar. The number of circles scattered on a blank 8 1/2" x 11" sheet was comparable to the original. However, due to settings on the computer program used to create the document, the size of the circles for the alternate version is approximately .10 mm larger than the original version. Therefore, there are 24 circles on the alternate forms while there are twenty-five circles on the original Trail Making B test. This almost imperceptible size and number difference was not viewed as detrimental to participants' test performance.

The original Trail Making B test requires participants to correctly connect the circles in ascending order, alternating numbers between letters; while the alternate test versions required that participants complete the task in the reverse order. The original Trail Making B task might not discriminate well enough between individuals who are functioning at the higher end of cognitive performance, so it was decided that participants would be required to complete the task in reverse order. The purpose of these "harder" test versions was to reduce potential ceiling effects yet, still tap similar cognitive processes as those assessed by the original Trail Making B test. Despite the similar layout of the circles on the sheet, participants begin with the number 12 instead of 1 and end with the letter A instead of the number 13. That is, while the original version requires a participant to draw lines from number 1 to letter A, 2 to B, 3 to C and so on; the alternate version begins with number 12 to letter L, 11 to K, 10 to J, 9 to I, 8 to H, 7 to G, 6 to F, 5 to E, 4 to D, 3 to C, 2 to B, and 1 to A. Similarly, the practice sheet contains the same number of practice items as the original, but begins with the number 12 to letter L, 11 to K, and so on.

Taking into account the preliminary research on pilot nutrition, the planned simulator study, and literature on cognitive performance tests, this study attempted to investigate the relationship between participants' scores on both the original and alternate versions of these constructed tests. Establishing the reliability of these alternate forms provided the basis for this study.

Method

Participants

Forty San Jose State University undergraduate students served as the participants in this study. Native and non-native English speaking male and female participants (32 women and 8 men, age range 18–30 years, mean age = 19.47 years) were recruited through sign-up sheets posted at the university. The sample consisted of 24 non-Native English speakers and 16 native English speakers. Seventeen students identified themselves as Asian American, 10 as Latino students, 6 as Caucasian students, 1 as an

African-American student, and 6 students chose the "Other" category. Exactly half of the sample indicated that they wore glasses to correct their vision. Thirty-eight reported no history of brain injury (2 had experienced minor concussions) and none of the participants had learning disabilities. There were 36 right-handers, 2 left-handers, and 2 ambidextrous participants. Inclusion in the study was voluntary and participants received 2 hours of research credit towards their Introduction to Psychology class.

Apparatus

The Wechsler Adult Intelligence Test–Third Edition (WAIS-III) was purchased through The Psychological Corporation (Wechsler, 1997). Subsections from the WAIS-III, specifically Digit Symbol-Coding and Letter-Number Sequencing were used in this study. The Trail Making B test (from the Halstead–Reitan Neuropsychological Battery) was purchased through the Reitan Neuropsychology Laboratory (Reitan, 1992). Three alternative versions of the Digit Symbol-Coding and Letter-Number Sequencing tests, and four alternative versions of the Trail Making B tests were created for this study. An electronic stopwatch, pencils, clipboard, testing sheets, television, and video recorder were utilized in a standard laboratory room.

Due to the expressed concerns of the copyright holders, because of the potential dissemination of the original tests and alternate versions of the DSC and LNS subtests, and the Trail Making B test to the public through this manuscript, none of these forms were included in the appendices. While the use of this material is solely for nonprofit educational purposes, it is in the best interests of the researchers for this information to remain contained. Original subtests from the WAIS-III are available and can be purchased through The Psychological Corporation. Trail Making B Test forms for Adults can be purchased from the Reitan Neuropsychology Laboratory.

In order to examine the usefulness of alternate forms, demonstrating the validity and reliability of the original tests is necessary. Concurrent validity of the WAIS-III is based on its correlation with several other measures (Wechsler, 1997). The new WAIS-III edition was correlated with the earlier WAIS-R (Revised) edition. The magnitude of the correlations among Verbal IQ scores (VIQ) = .94, Performance IQ scores (PIQ) = .86, and Full Scale IQ scores (FSIQ) = .93, suggest that the WAIS-III measures the same constructs as the WAIS-R. In addition, the WAIS-III was correlated with the WISC-III (Wechsler Intelligence Test for Children–Third Edition) and showed that it was an extension of the respective children's edition. The tests correlations were high and statistically significant, VIQ = .88, PIQ = .78, and FSIQ = .88. Correlations between the WAIS-III and the WIAT (Wechsler Individual Achievement Test) provide evidence of the validity of these tests. The composite scores (VIQ, PIQ, and FSIQ) are high, between the .60s and .70s and range between .53 and .81. The correlations between the WIAT composite scores and the WAIS-III's PSI (or Processing Speed Index Score) range from .45 to .65 with most between .55 and .65. Correlations were also conducted between the WAIS-III (not including the Letter-Number Sequencing subtest) and the Stanford Binet Intelligence Scale–Fourth Edition (SB-IV) and found that the global SB-IV composite score is .88, which is consistent with studies that have tested the relationship between the Wechsler scales and the SB-IV. These results also suggest that the WAIS-III has a strong relationship with the SB-IV, as did the WAIS-R. In addition, very consistent results were found with predecessors of the WAIS-III (not including the Letter-Number Sequencing subtest) and the Standard Progressive Matrices, with correlations from the .50s to .70s.

The WAIS-III is also correlated with several other measures of "various domains of functioning including overall cognitive ability, memory, attention and concentration, fine motor speed, fine motor dexterity, spatial processing, executive functioning, and language" (Wechsler, 1997). Correlation studies between the WAIS-III and the WAIS-R, the WISC-III, the WIAT, the SBI-IV, and the Standard Progressive Matrices provide evidence for its concurrent validity. The WAIS-III was also administered to individuals with mental retardation, learning disabilities, hearing impairments, psychiatric disorders, and neuropsychological deficits (e.g., Alzheimer's dementia, traumatic brain injury) in order to provide evidence for its construct validity and clinical utility (Wechsler, 1997).

All WAIS-III subtests were found to significantly correlate with each other. Subtests that assess similar functioning correlated more highly with each other than with tests measuring different types of functioning. The Digit Symbol-Coding and Letter-Number Sequencing subtests correlate at .44.

Exploratory factor analyses were then conducted on subtests of the WAIS-III and patterns loaded into four factors, which were identified as structures for discrete units of functioning (i.e., First Factor: Verbal Comprehension, Second Factor: Perceptual Organization, Third Factor: Working Memory, and Fourth Factor: Processing Speed). The pattern of the factor solutions revealed that Letter-Number Sequencing loads highly onto Working Memory (.62), while Digit Symbol-Coding loads onto Processing Speed (.68). This four-factor model has empirical support across the age ranges (16-89 years).

In order to establish reliability of the Letter-Number Sequencing WAIS-III subtest, split-half reliability coefficients corrected by the Spearman-Brown formula were conducted (Wechsler, 1997). Since Digit Symbol-Coding is a speeded subtest, the split-half coefficient is not an appropriate method of estimating reliability. Instead, test-retest coefficients were used to establish reliability. The average (across age groups 16–89 years and calculated with Fisher's z transformation) reliability coefficient for the Digit Symbol-Coding subtest is .84, while the average reliability coefficient for the Letter-Number Sequencing subtest is .82.

The Trail Making B test "is a commonly used neuropsychological instrument that has demonstrated validity and reliability…it is proven to be sensitive to the effects of cerebral dysfunction, and its clinical utility is well-established" (Breece, 1998, p. 56). Factor analytic investigations have not delineated all of the cognitive abilities that the TMB test may measure (Groff & Hubble, 1981). Discussion has addressed whether the Trail Making B assesses a specific ability or several underlying abilities. However, many studies have verified good construct validity and have determined that the Trail Making B test clearly assesses some aspects of mental processing speed, mental flexibility, visual scanning, motor functioning, and attention (Corrigan & Hinkeldey, 1987; Lezak, 1995; O'Donnell, MacGregor, Dabrowski, Oestreicher, & Romero, 1994).

Reitan (1959) computed Pearson product-moment coefficients of correlation between the Trail Making Test and the Wechsler–Bellevue Scale (Form I), transformed them to Fisher's z values, and found significant correlations between variables – .70 for certain subtests. High inter-correlations between the Trail Making B test and the

Wechsler–Bellevue Verbal IQ (.60), Performance IQ (.71), and Full Scale IQ (.70) variables exist among non-brain damaged participants. These correlations favorably compare with inter-correlations of Wechsler–Bellevue subtests found by Wechsler (1944). Fitzhugh, Fitzhugh, and Reitan (1962) also found significant coefficients of correlations between the Trail Making test and Wechsler–Bellevue Verbal IQ (.66), Performance (.53), and Full IQ (.77) variables, for the control group. The Digit Symbol-Coding subtest correlated significantly with the Trail Making B test (.82) at the $p < .001$ level.

Design

The variables in this study were the participants' scores on the original and alternate versions of the tests. Participant's scores on the original and alternative tests were examined within each test type (Digit Symbol-Coding, Letter-Number Sequencing, Trail Making B) using alternate forms reliability, or the correlation coefficient of equivalence (r). "A correlation between the scores on each form indicates the alternate forms reliability, also called the coefficient of equivalence. The higher the correlation between the two sets of scores the more equivalent the forms are considered" ("General and Specific Issues", 1996, p. 6). Scores on the alternate forms of the Trail Making B test were expected to be lower than those of the original form since the task required participants to complete the task (i.e., shifting between the alphabet and numbers) in reverse to make the task more difficult, although still tapping relatively the same processes. In order to correctly retain the null hypothesis (while controlling for Type

II error), that there were no significant differences among the alternate test forms, a series of power analyses was performed to estimate sample sizes (Keppel, Saufley, & Tokunaga, 1992; Cohen, 1988). Since these are alternate forms of pre-existing tests the power was set at .80 ($\delta = 2.8$).

Significant positive correlations among the versions were required to attain sufficient reliability. An alpha level of 0.05 ($\alpha = .05$) was set for all reliability coefficients. The difference between a reliability coefficient of 0.7 and 0.9 would have required 25 participants to achieve power of .80; while the difference between a reliability coefficient of 0.8 and 0.9 would have required 60 participants (Uitenbroek, 2001). Therefore, in order to detect the expected minimal differences between test versions and maintain an acceptable or "good enough" level of difference between the tests for this study, the sample size included 40 participants. Researchers expected that by setting high power (.80; Cohen, 1988), if differences did exist between test versions this conservative approach would detect these minimal differences.

Procedure

Participants voluntarily signed a consent form (see Appendix B) and completed a brief demographics sheet (see Appendix C). Participants were assigned an identification number and scheduled for two meeting periods, set approximately one day apart. Individual sessions were scheduled to prevent shortcuts, to encourage following directions, and to adhere to the original test design of the DSC subtest, LNS subtest, and TMB test as being individually administered tests. These tests were developed for face-to-face administration. Participants completed one of four pre-set protocols during each testing session. Each protocol contained one version of each of the following tests. There were (timed) paper and pencil tests for Digit Symbol-Coding and Trail Making B and the experimenter verbally administered the (un-timed) Letter-Number Sequencing subtest.

During each of the two meeting periods, participants completed one set of each of the three tests and then, took a 15-20 minute break while watching a "neutral" wildlife documentary videotape (Jones, 1991). After the break, the experimenter returned to the testing room and administered a second set of each of the three tests. Thus, all participants completed all four versions of the three tests across the two testing sessions. During the first initial session, all participants also completed the original Trail Making B test.

The investigator familiarized participants with the tests and testing requirements. The test version packets (protocol) and the order of presentation of the tests were organized in consideration of possible combinations of tests, as well as order (i.e., Latin square design). The instructions given to the participants were those specifically provided in the original WAIS-III and Trail Making B testing kits. Raw scores were converted into scaled scores as required for the Digit Symbol-Coding and Letter-Number Sequencing subtests (see Appendix D). The testing time was approximately forty-five minutes per session.

The test and re-test reliability of scoring procedures has received some criticism (Lezak, 1982), as it is the test examiner's duty that "if the subject makes an error, point it out immediately, return the subject to the last correct circle, and continue the test from that point. Do not stop timing" (Reitan, 1992, p. 5). However, alternative methods of interpreting test scores have been proposed (Corrigan & Hinkeldey, 1987; Lamberty et al., 1994). These methods involve creating a ratio score using the more basic Trail Making A test, where participants correctly connect numbers only (from 1 to 25) in ascending order, in combination with the Trail Making B score. Research continues in interpretation of Trail Making B test scores. Therefore, Reitan's (1992) most commonly used scoring method for the Trail Making B test was employed in this study.

Results

A table that describes the means and standard deviations of participants' scores on each of the tests and versions is included (see Table 3).

To investigate the reliability of the alternate forms, Pearson product-moment correlation coefficients (r) were computed for each version within each test type at the .05 level. All four versions of the Digit Symbol-Coding subtests were significantly and positively correlated (range .74 to .92) (see Table 4). All of the four versions of the Letter-Number Sequencing subtests were also significantly and positively correlated (range .54 to .82) (see Table 5). Significant and non-significant correlations were found among the alternate versions of the Trail Making B (TMB) test (range .08 to .59) (see Table 6). A table that describes the strength of the relationships and proportions of shared variability between the test versions was included (see Table 7). All versions of the Digit Symbol-Coding and Letter-Number Sequencing subtests exhibited "large" to "extremely large" relationships. For example, versions 1 and 2 of the Digit Symbol-Coding subtest shared a proportion of 76% variance. In other words, 76% of the

Table 3

Scores on the DSC, LNS, and TMB by Test Version

	M	SD
Digit Symbol-Coding		
Version 1 (original)	12.65	2.43
Version 2	12.28	2.69
Version 3	12.08	2.55
Version 4	11.28	2.56
Letter-Number Sequencing		
Version 1 (original)	10.30	2.29
Version 2	9.75	2.25
Version 3	10.15	2.44
Version 4	10.07	2.37
Trail Making B		
Original Version	64.98	16.24
Version 1	71.57	25.68
Version 2	64.13	23.02
Version 3	83.35	27.17
Version 4	83.05	29.47

Note. The values for DSC and LNS represent means of scaled scores. The values for TMB represent completion time in seconds. n = 40 per each group.

Table 4

Intercorrelations of DSC Scores Across Test Versions

Version	1	2	3	4
		Participants (n = 40)		
Version 1	—	.87*	.77*	.74*
Version 2		—	.80*	.79*
Version 3			—	.92*
Version 4				—

Note. * $p < .05$.

Table 5

Intercorrelations of LNS Scores Across Test Versions

Version	1	2	3	4
		Participants (n = 40)		
Version 1	—	.58*	.60*	.63*
Version 2		—	.54*	.59*
Version 3			—	.82*
Version 4				—

Note. * $p < .05$.

Table 6

Intercorrelations of TMB Scores Across Test Versions

Version	Original	1	2	3	4
		Participants (n = 40)			
Original Version	—	.48*	.41*	.08	.18
Version 1		—	.59*	.16	.25
Version 2			—	.23	.34*
Version 3				—	.53*
Version 4					—

Note. * $p < .05$.

Table 7

Correlation as a Function of Strength & Residual Variation

Test & Version	r	r^2	Residual Variation
Digit Symbol-Coding			
Version 1 & 2	.87	.76	.24
Version 1 & 3	.77	.59	.41
Version 1 & 4	.74	.55	.45
Version 2 & 3	.80	.64	.36
Version 2 & 4	.79	.62	.38
Version 3 & 4	.92	.85	.15
Letter-Number Sequencing			
Version 1 & 2	.58	.34	.66
Version 1 & 3	.60	.36	.64
Version 1 & 4	.63	.40	.60
Version 2 & 3	.54	.29	.71
Version 2 & 4	.59	.35	.65
Version 3 & 4	.82	.67	.33
Trail Making B Test			
Original & 1	.48	.23	.77
Original & 2	.41	.17	.83
Original & 3	.08	.01	.99
Original & 4	.18	.03	.97
Version 1 & 2	.59	.35	.65
Version 1 & 3	.16	.03	.97
Version 1 & 4	.25	.06	.94
Version 2 & 4	.34	.12	.88
Version 3 & 4	.53	.28	.72

Note. Cohen (Keppel et al., 1992) = "small" relationship as $r = +/- .10$, or $r^2 = .01$
"medium" relationship as $r = +/- .30$, or $r^2 = .09$
"large" relationship as $r = +/- .50$, or $r^2 = .25$

variability on version 1 is explained or can be accounted for by performance on version 2. Similarly, 24% of the variability in scores on version 1 is unrelated or unaccounted for residual variation to scores on version 2.

Separate repeated measures of analyses of variance (ANOVA) were computed for each test type among the different test versions. Comparisons were then performed to determine which versions differed significantly from one another. A 4 (test version) x 4 (test order) mixed factorial analysis was also conducted in order to examine the effect of test version by test order. Test order was the between subjects factor.

The analyses indicated that there was a significant effect of test version among the original and alternate forms of the Digit Symbol-Coding subtest, $F\ (3, 39) = 12.22$, $p < .05$. Comparisons showed that versions 1 ($M = 12.65$) and 2 ($M = 12.28$) did not significantly differ, $F\ (1, 39) = 3.26, p > 1$; nor did versions 2 ($M = 12.28$) and 3 ($M = 12.08$), $F\ (1, 39) = .58, p > 1$. Versions 3 ($M = 12.08$) and 4 ($M = 11, 28$) did significantly differ, $F\ (1, 39) = 23.55, p < .05$; versions 2 ($M = 12. 28$) and 4 ($M = 11.28$) significantly differed, $F\ (1, 39) = 13.93, p < .05$; versions 3 ($M = 12.08$) and 1 ($M = 12.65$) significantly differed, $F\ (1, 39) = 4.53, p < .05$; and versions 4 ($M = 11.28$) and 1 ($M = 12.65$) significantly differed, $F\ (1, 39) = 22.80, p < .05$. A significant interaction effect was found between the effect of test version by order for the original and alternate forms of the Digit Symbol-Coding subtests, $F\ (9, 39) = 2.44, p < .05$; however, a main effect for order of presentation was not found, $F\ (3, 36) = .74, p > 1$ (see Figure 1).

A significant effect of test version did not exist within the original and alternate forms of the Letter-Number Sequencing subtest, $F\ (3, 39) = 1.05, p > 1$. There was no significant interaction between the effect of test version by order for the original and alternate forms of the Letter-Number Sequencing subtests, $F(9, 39) = 1.03$, $p > 1$, and a main effect for order of presentation was not found, $F(3, 36) = .50$, $p > 1$ (see Figure 2).

A significant effect of version was found within the alternate versions of the Trail Making B test, $F(3, 39) = 9.91$, $p < .05$. Comparisons showed that versions 1 (M = 71.57) and 2 (M = 64.13) significantly differed, $F(1,39) = 4.51$, $p < .05$; versions 1 (M = 71.57) and 3 (M = 83.35) significantly differed, $F(1, 39) = 4.72$, $p < .05$; versions 1 (M = 71.57) and 4 (M = 83.05) significantly differed, $F(1, 39) = 4.58$, $p < .05$; versions 2 (M = 64.13) and 3 (M = 83.35) significantly differed, $F(1, 39) = 15.06$, $p < .05$; and versions 2 (M = 64.13) and 4 (M = 83.05) significantly differed, $F(1, 39) = 15.28$, $p < .05$. Alternate versions 3 (M = 83.35) and 4 (M = 83.05) did not significantly differ, $F(1, 39) = .01$, $p > 1$.

The original Trail Making test version (M = 64.98) did not significantly differ from alternate version 1 (M = 71.57), $F(1, 39) = 3.32$, $p > 1$ or version 2 (M = 64.13), $F(1, 39) = .06$, $p > 1$. The original version (M = 64.98) did significantly differ from version 3 (M = 83.35), $F(1, 39) = 14.49$, $p < .01$ and version 4 (M = 83.05), $F(1, 39) = 13.55$, $p < .05$. A significant interaction was found between the alternate forms of the Trail Making B test and order of presentation, $F(9, 39) = 4.79$, $p < .01$; but a main effect for order of presentation was not found, $F(3, 36) = .79$, $p > 1$ (see Figure 3).

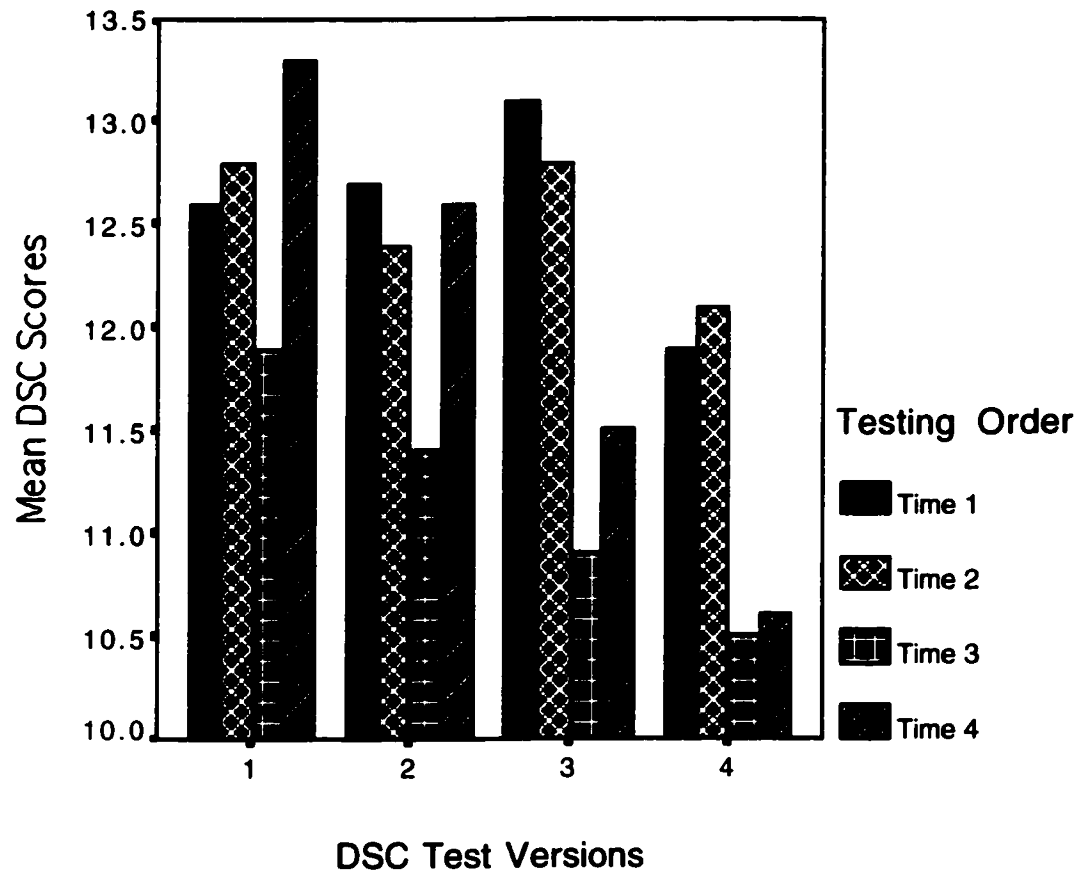

Figure 1. Mean scores on the Digit Symbol-Coding subtest as a function of version and order (n = 40). Version 1 = Original Version. Versions 2, 3, & 4 = Alternate forms.

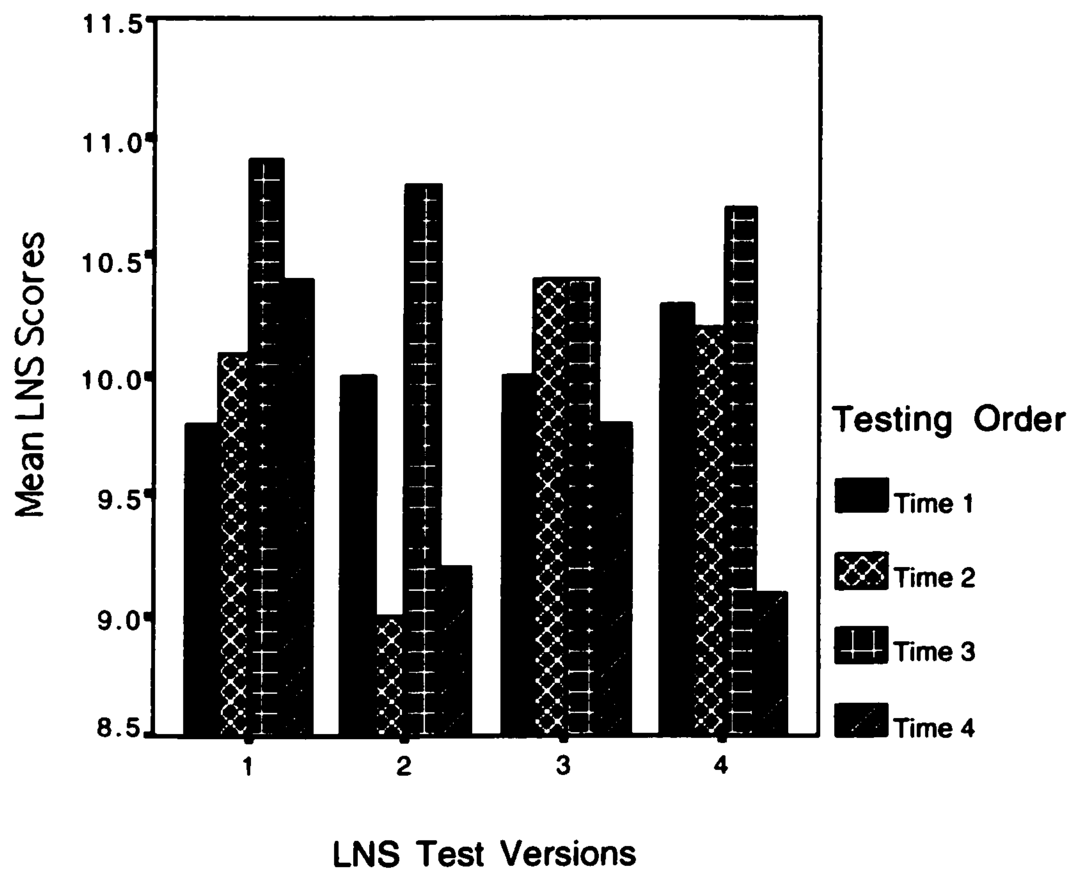

Figure 2. Mean scores on the Letter-Number Sequencing subtest as a function of version and order (n = 40). Version 1 = Original Version. Versions 2, 3, & 4 = Alternate forms.

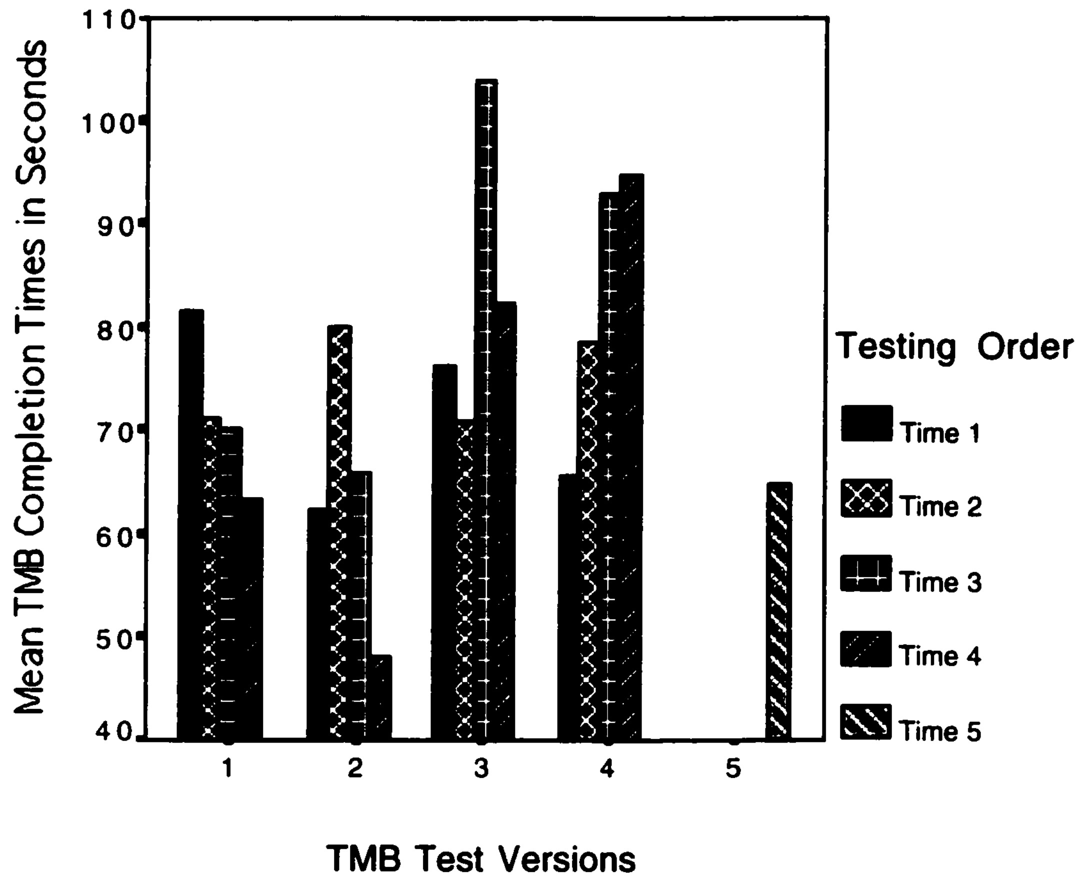

Figure 3. Mean scores on the Trail Making B test as a function of version and order (n = 40). Version 5 = Original Version, Time 5 = First initial session.

Discussion

The reliability analyses confirmed moderate to strong significant correlations among all of the original and alternate versions of the Digit Symbol-Coding and Letter-Number Sequencing subtests. Significant and non-significant moderate to weak correlations were found among the original and alternate versions of the Trail Making B tests. It would not have been surprising to find practice effects due to repeated administrations; in this study these effects were non-significant.

Alternate forms reliability was established when the test scores were significantly correlated, with strong relationships within the different tests. A main effect of version existed for the alternate forms of the Digit Symbol-Coding subtest and Trail Making B test, but not for the Letter-Number Sequencing subtest. An effect of version suggests that the mean difference of the sets of scores from participants varied in range within the versions. That is, participants generated higher scores on some versions than on others, but were consistent with this performance across the versions. This effect indicates caution in assigning the different versions to conditions. In subsequent studies, all testing orders and versions should be equally represented in each testing condition. It would also be beneficial to counterbalance the practice tests in order to control for any effect of pre-test priming effects.

Analyses indicate sufficiently acceptable levels of reliability for the original and alternate versions of the Digit Symbol-Coding, and Letter-Number Sequencing subtests. However, this level was not reached for the alternate versions of the Trail Making B test. Again, with the caveat that a method of counterbalancing test version to condition be employed, use of the alternate versions of the Digit Symbol-Coding and Letter-Number Sequencing subtests can be implemented.

In explaining the reliability of these alternate forms, we review Rudner and Schafer's (2000) discussion on reliability. The principal issue in reliability analysis concerns nonsystematic or random errors. An "observed score" is comprised of a participant's "true score", "systematic error", and "random error". As the reliability coefficient approaches 1.0, more of the variability in the observed score is due to true score variability rather than random error.

Rudner and Schafer (2000) describe that two of the three major sources of testing error include factors in the test itself and factors relating to the participants taking the test. These two sources were contributing factors in this study. The third possible source of error –or factors related to scoring– was minimized; the scoring procedure was fairly simple and standardized. Issues related to test administration may have played a role, and will be discussed later. The error related to factors in the test itself is contributed to by the complexity of the domains; that is, the more complex the domain, the more difficult it is to create items which clearly reflect or test that domain (e.g., arithmetic skills versus political attitudes). This may have occurred in the alternate forms of the Trail Making B test. The complexities of not only the layout, but of the cognitive faculties pertinent to completing the task were not apparent.

The lack of heterogeneity in the sampled population can be viewed as a potential weakness in this study. Rudner and Schafer (2000) assert that, "reliability is a joint characteristic of a test and examinee group, not just a characteristic of a test" (p. 3).

This homogeneous population, of similar age and education, may have unduly increased the reliability of the measures. While assessing reliability is essentially assessing consistency, it was found that despite the variability of the obtained scores, two of the three measures reached acceptable levels of reliability.

The failure to find reliable alternate versions of the Trail Making B test limits the ability to utilize these alternate forms for subsequent research. The method of alternate test version construction for this test was the least controlled. Namely, adherence to the Trail Making B test's layout was not maintained. Despite alterations to the circle size diameter and number of circles on the page (from 25 to 24), the correlations among the four alternate versions were weak and non-significant. For the Digit Symbol-Coding and Letter-Number Sequencing subtests, alternate forms construction involved less manipulation in re-numbering and re-alphabetizing items. This process differed from the construction of the Trail Making B test, where it required a systematic method of placing numbers and letters in circles on a blank page in a specific manner. It may prove beneficial in future test construction of Trail Making B alternate forms to use rotated mirror images of the layout to create alternate forms. The identical distances and patterns among items may reduce the variance among the versions. Unfortunately, the inconsistency in these alternate test versions was highly susceptible to differences in test layout and pathway construction, thus negatively affecting reliability.

An additional issue regarding the Trail Making B test and alternate versions involves test administration. While there has been criticism about the effect of the reaction time of the examiner to point out mistakes (Lezak, 1995), one recommendation for administration of alternate forms is that the examiner should present the blank side of the form and require participants to turn over the form with the items, at which point s/he begins timing. While these are the instructions for the original Trail Making B, this was not carried out in this study and may have given some participants an unfair advantage. This suggested method might better control for participants who scan the sheet before timing begins, versus those who do not, and might allow for a "truer" performance score.

The alternate forms of the Trail Making B test were expected to be "harder" due to performing the task in reverse, possibly affecting visual scanning and increasing testing layout complexity as found in other alternate form studies of the Trail Making B test (Breece, 1998). A main effect of version and subsequent test version comparisons did indicate that the complexity of the task differed between the original and alternate versions. Participants took significantly longer to complete versions 3 and 4 compared to the original version. However, alternate versions 1 and 2 correlated significantly with the original version establishing alternate forms reliability, while versions 3 and 4 did not. Regardless of the task being "harder" (resulting in significantly longer completion times) in some cases, it is difficult to determine why exactly participants performed more poorly and more inconsistently as determined by correlation coefficients. It may be possible that the very nature of the task was altered from version to version, and/or resulting from other discrete factors. It may be more useful to create alternate forms, or rotated mirror images based upon versions 1 and 2, since these versions exhibited consistency with moderate significant correlations to the original version and a "large" relationship of correlational strength. Some levels of correlational relationships did exist among the alternate versions. However, since the alternate forms of the Trail Making B as a group were not sufficiently reliable, these test versions should not be implemented as pre- and post-test measures.

A single investigator collected all data from participants, reducing discrepancies during the timed-tests, and controlling for possible administration effects. The researcher also attempted to control for practice and ordering effects by counterbalancing the different test versions and allowing two separate sessions, with a break between sessions. However, issues such as fatigue, lack of personal interest in performance on both timed and un-timed tests, emotional distractions, and background noises can explain some of the sources of variance in individuals' performance.

While language difficulties could also explain possible differences between performance of native and non-native English speakers, this was not within the scope of this study nor considered in the analyses. It was evident that had language difficulties been an issue, it would not affect the consistency of performance among the different versions; that is, if a participant performed poorly on one version of the test due to language, they would perform poorly on all versions and vice versa. Thus, a clear understanding of the directions was encouraged.

Conclusion

The present study attempted to create reliable alternate forms of tests that assess various cognitive functions. The goal was to implement the alternate forms in subsequent aviation related repeated measures studies. Overall, this endeavor proved successful with the alternate test versions of the Digit Symbol-Coding and Letter-Number Sequencing subtests exhibiting strong reliability. These brief, reliable cognitive measures can be useful in the aviation operational environment, and will allow for investigation into factors affecting working memory, ability to shift mental sets, cognitive processing speed, distractibility, and other cognitive faculties.

Perhaps a lack of methodological control in test construction of the Trail Making B test resulted in insufficiently reliable and unequal versions. Suggestions for construction of alternate versions of this specific test in the future have been addressed. Gleaning the importance of the discrete differences that can affect cognitive test development and construction, it has been demonstrated that measurement design requires strict adherence to, and a clear understanding of, a test's distinct underlying psychometric properties.

References

Anastasi, A. (1988). *Psychological Testing* (6th ed., pp. 109-120). New York: Macmillan Publishing Company.

Aviation Safety Reporting System. (2001a). [Personal Database Search]. Retrieved January 4, 2001, from http://nasdac.faa.gov/asp/asy_asrs.asp.

Aviation Safety Reporting System. (2001b). Search request No. 6110, *Hypoglycemia Reports* [Database Search]. ASRS Office, Mountain View, California: Battelle.

Billings, C.E., & Reynard, W.D. (1984). Human Factors in Aircraft Incidents: Results of a 7-Year Study. *Aviation, Space, and Environmental Medicine*, October: 960-965.

Bischoff, L.G., Warzak, W.J., Maguire, K.B., & Corely, K.P. (1992). Acute and Chronic Effects of Hypoglycemia on Cognitive and Psychomotor Performance. *Nebraska Medical Journal, 77* (9), 253-262.

Blackman, J.D., Towle, V.L., Lewis, G.F., Spire, J.P., & Polonsky, K.S. (1990). Hypoglycemic Thresholds for Cognitive Dysfunction in Humans. *Diabetes, 39*, 828-835.

Breece, B.A. (1998). *Test-Retest Reliability of the Expanded Trail Making Test and Its Alternate Forms*. Unpublished doctoral dissertation, California School of Professional Psychology, Fresno, CA.

Canfield, D.V., Chaturvedi, A.K., Boren, H.K., Véronneau, S.J.H., & White, V.L. (2000). *Abnormal Glucose Levels Found in Transportation Accidents*. (FAA Publication Report No. DOT/FAA/AM-00/22, pp. 1-11). Washington, D.C.: Office of Aviation Medicine.

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (2nd ed., pp. 75-83). New Jersey: Lawrence Erlbaum Associates.

Corrigan, J.D., & Hinkeldey, N.S. (1987). Relationship Between Parts A and B of the Trail Making Test. *Journal of Clinical Psychology, 43* (4), 402-409.

Cox, D.J., Gonder-Frederick, L., Antoun, B., Cryer, P.E., & Clarke, W.L. (1993). Perceived Symptoms in the Recognition of Hypoglycemia. *Diabetes Care, 16* (2), 519-527.

Cox, D.J., Gonder-Frederick, L.A., Kovatchev, B.P., Julian, D.M., & Clarke, W.L. (2000). Progressive Hypoglycemia's Impact on Driving Simulation Performance. *Diabetes Care, 23* (2), 163-170.

Cox, D.J., Gonder-Frederick, L.A., Schroeder, D.B., Cryer, P.E., & Clarke, W.L. (1993). Disruptive Effects of Acute Hypoglycemia on Speed of Cognitive and Motor Performance. *Diabetes Care, 16* (10), 1391-1393.

Dröge, C. (1996, September/October). How Valid Are Measurements? *Decision Line, 27* (5). Retrieved August 19, 2001, from http://www.decisionsciences.org/Newsletter/vol27/27_5/res27_5.htm.

Driesen, N.R., Cox, D.J., Gonder-Frederick, L., & Clarke, W. (1995). Reaction Time Impairment in Insulin-Dependent Diabetes: Task Complexity, Blood Glucose Levels, and Individual Differences. *Neuropsychology, 9* (2), 246-254.

Doraz, W.E. (1987). Diet and Delinquency: The Ground of Four Leading Theories in Human Physiology and Sociology. In W.B. Essman (Ed.), *Nutrients and Brain Functions* (pp. 222-223). Basel: Karger.

Evans, M.L., Pernet, A., Lomas, J., Jones, J., & Amiel, S.A. (2000). Delay in Onset of Awareness of Acute Hypoglycemia and of Restoration of Cognitive Performance During Recovery. *Diabetes Care, 23* (7), 893.

Farrace, S., Urbani, L., Sakara, L., & DeAngelis, C. (1993). *Idiopathic reactive hypoglycemic disorders in a population of healthy trainees of an Italian air force military school.* Paper presented at the AGARD Conference Proceeding 533: Nutrition, Metabolic Disorders, and Lifestyle of Aircrew, Oslo, Norway.

Field, J.B. (1989). Hypoglycemia: Definition, Clinical Presentations, Classification, and Laboratory Tests. *Endocrinology and Metabolism Clinics of North America, 18* (1), 27-43.

Fisher, M.G.P., & Atkinson, D.W. (1980). Fasting or feeding? A survey of fast-jet aircrew nutrition in the Royal Air Force Strike Command, 1979. *Aviation, Space, and Environmental Medicine*, October, 1119-1121.

Fitzhugh, K.B., Litzhugh, L.C., Reitan, R.M. (1962). Relation of Acuteness of Organic Brain Dysfunction to Trail Making Test Performances. *Perceptual and Motor Skills, 15*, 399-403.

Ford, C.N., Jack, T.D., Crisp, V., & Sandusky, R. (1999). *Aviation Accident Causal Analysis*. Paper presented at the Advances in Aviation Safety Conference and Exposition, Daytona Beach, FL.

General and Specific Issues. (n.d.). Retrieved July, 16, 2001, from http://nces.ed.gov/npec/evaltests/references.html.

Gibbons, H.L., Plechus, J.L., Chandler, E.H., & Ellis, J.W. (1966). Alcohol-induced Hypoglycemia as a Factor in Aircraft Accidents. *Aerospace Medicine, 37* (9), 959-961.

Gold, A. E., Deary, I.J., MacLeod, K.M., & Frier, B.M. (1995a). The effect of IQ level on the degree of cognitive deterioration experienced during acute hypoglycemia in normal humans. *Intelligence, 20*, 267-290.

Gold, A.E., Deary, I.J., MacLeod, K.M., Thomson, K.J., & Frier, B.M. (1995b). Cognitive function during insulin-induced hypoglycemia in humans: short-term cerebral adaptation does not occur. *Psychopharmacology, 119*, 325-333.

Gold, A.E., MacLeod, K.M., Deary, I.J., & Frier, B.M. (1995c). Hypoglycemia-Induced Cognitive Dysfunction in Diabetes Mellitus: Effect of Hypoglycemia Unawareness. *Physiology & Behavior, 58*, (3), 501-511.

Gonder-Frederick, L.A., Cox, D.J., Driesen, N.R., Ryan, C.M., & Clarke, W.L. (1994). Individual Differences in Neurobehavioral Disruption During Mild and Moderate Hypoglycemia in Adults With IDDM. *Diabetes, 43*, 1407-1412.

Gregory, R.J. (1999). *Foundations of Intellectual Assessment: The WAIS-III and Other Tests in Clinical Practice*. Boston, MA: Allyn and Bacon.

Grimm, L.G., Yarnold, P.R. (2000). *Reading and Understanding More Multivariate Statistics* (pp. 23-31, 99-117). Washington, D.C.: American Psychological Association.

Groff, M.G., & Hubble, L.M. (1981). A Factor Analytic Investigation of the Trail Making Test. *Clinical Neuropsychology, 3* (4), 11-13.

Guohua, L. (1994). Pilot-Related Factors in Aircraft Crashes: A Review of Epidemiologic Studies. *Aviation, Space, and Environmental Medicine,* (October), 944-952.

Harper, C.R., & Kidera, G.J. (1973). Hypoglycemia in Airline Pilots. *Clinical Aviation and Aerospace Medicine,* July, 769-771.

Hoffman, R.G., Speelman, D.J., Hinnen, D.A., Conley, K.L., Guthrie, R.A., Knapp, R.K. (1989). Changes in Cortical Functioning with Acute Hypoglycemia and Hyperglycemia in Type I Diabetes. *Diabetes Care, 12* (3), 193-197.

Holmes, C.S., Koepke, K.M., & Thompson, R.G. (1986). Simple Versus Complex Performance Impairments at Three Blood Glucose Levels. *Psychoneuroendocrinology, 11* (3), 353-357.

Introduction to Reliability and Validity. (n.d.). Retrieved August 19, 2001, from http://courses.wcupa.edu/bolton/reliabilityvalidity/reliability.htm.

Jones, P. (Executive Producer). (1991). *David Attenborough's: The Trials of Life - Finding Food* [Time Life Video Series]. Alexandria, VA: Time Life, Inc.

Kelland, D.Z., & Lewis, R.F. (1994). Evaluation of the Reliability and Validity of the Repeatable Cognitive-Perceptual-Motor Battery. *The Clinical Neuropsychologist, 8* (3), 295-308.

Keppel, G., Saufley, W.H. Jr., Tokunaga, H. (1992). *Introduction to Design and Analysis: A Student's Handbook* (2nd ed.). New York: W.H. Freeman and Company.

Lamberty, G.J., Putnam, S.H., Chatel, D.M., Bieliauskas, L.A., & Adams, K.M. (1994). *Neuropsychiatry, Neuropsychology, and Behavioral Neurology, 7* (3), 230-234.

Lezak, M.D. (1995). *Neuropsychological assessment* (3rd ed.). New York: Oxford University Press.

Liungman, Carl G. (1991). *Dictionary of Symbols*. New York, NY: W.W. Norton & Company.

McCrimmon, R.J., Deary, I.J., & Frier, B.M. (1997). Auditory information processing during acute insulin-induced hypoglycaemia in non-diabetic human subjects. *Neuropsychologia, 35* (12), 1547-1553.

McCrimmon, R.J., Deary, I.J., Huntly, J.P., MacLeod, K.J., & Frier, B.M. (1996). Visual information processing during controlled hypoglycaemia in humans. *Brain, 119*, 1277-1287.

Merck Manual Home Edition. (2001). *Hypoglycemia* [Electronic version]. Retrieved January 31, 2001, from http://www.merck.com/pubs/mmanual_home/sec13/148.htm.

Meyer, J.F. (1969). Blood Glucose During High-performance Aircraft Flight. *Aerospace Medicine* (March), 310-315.

National Transportation and Safety Board. (1979). Aircraft Accident Report. *Air New England, Inc. deHavilland DHC-6-300, N383EX, Flight 248, Hyannis, Massachusetts, June 17, 1979* (Report No. NTSB-AAR-80-1). Washington, D.C.: Author.

National Transportation and Safety Board. (1976). Aircraft Accident Report. *Alaska Airlines, Inc. Boeing 727-81, N124AS, Flight 60, Ketchikan International Airport, Ketchikan, Alaska, April 5, 1976* (Report No. NTSB-AAR-76-24). Washington, D.C.: Author.

National Transportation and Safety Board. (1997). Aircraft Accident Report. *Beech G35, N4552D, Jackson, Wyoming, July, 6, 1997* (Report No. NTSB-SEA97LA159). Washington, D.C.: Author.

National Transportation and Safety Board. (1989). Aircraft Accident Report. *Cessna 150, N704SM, Orlando International Airport, Orlando, Florida, December 26, 1989* (Report No. NTSB-MIA-90-L-A041). Washington, D.C.: Author.

National Transportation and Safety Board. (1998). Aircraft Accident Report. *Extra Flugzeugbau EA-300, N301NL, Chenoa, Illinois, April 12, 1998* (Report No. NTSB-CHI98LA121. Washington, D.C.: Author.

National Transportation and Safety Board. (1992). Aircraft Accident Report. *Luscombe 8A, N71183, Snoqualmie, Washington, August 24, 1992* (Report No. NTSB-SEA92LA188). Washington, D.C.: Author.

National Transportation and Safety Board. (1986). Aircraft Accident Report. *Pan American, Boeing 747-121, N751PA, Flight 362, Miami, Florida, December 21, 1986.* (Report No. NTSBMIA87IA054). Washington, D.C.: Author.

O'Donnell, J.P., MacGregor, L.A., Dabrowski, J.J., Oestreicher, J.M., & Romero, J.J. (1994). Construct Validity of Neuropsychological Tests of Conceptual and Attentional Abilities. *Journal of Clinical Psychology, 50*, 4, 596-600.

Porcu, S., Berti, R., & Lala, A. (1987). *Spontaneous cerebral electrical activity during prolonged hypoglycemia.* Paper presented at the AGARD Conference Proceeding 432: Electric and Magnetic Activity of the Central Nervous System: Research and Clinical Applications in Aerospace Medicine, Trondheim, Norway.

Powell, T.J. (1956). Episodic Unconsciousness in Pilots during Flights. *Journal of Aviation Medicine, 27,* 301-316.

Powell, T.J., Carey, T.M., Brent, H.P., & Taylor, W.J.R. (1957). Episodes of Unconsciousness in Pilots during Flight in 1956. *Journal of Aviation Medicine, 28,* 374-386.

Raichle, M.E., & King, W.H. (1972). Functional Hypoglycemia: A Potential Cause of Unconsciousness in Flight. *Aerospace Medicine,* (January).

Reitan, R.M. (1959). Correlations Between the Trail Making Test and the Wechsler-Bellevue Scale. *Perceptual and Motor Skills, 9,* 127-130.

Reitan, R.M. (1992). *Trail Making Test: Manual for Administration and Scoring.* Tucson: Reitan Neuropsychology Laboratory.

Rudner, L.M., & Schafer, W.D. (2001). *Reliability* (ERIC Identifier No. ED458213). Retrieved May 30, 2002, from ERIC Clearinghouse on Assessment and Evaluation, ERIC database website: http://www.ed.gov/database/ERIC_Digests/ed458213.html.

Stanczak, D.E., Lynch, M.D., McNeil, C.K., & Braun, B. (1998). The Expanded Trail Making Test: Rationale, Development, and Psychometric Properties. *Archives of Clinical Neuropsychology, 13* (5), 473-487.

Stevens, A.B., McKane, W.R., Bell, P.M., Bell, P., King, D.J., & Hayes, J.R. (1989). Psychomotor Performance and Counterregulatory Responses During Mild Hypoglycemia in Healthy Volunteers. *Diabetes Care, 12* (1), 12-17.

Tabandeh, H., Ranganath, L., & Marks, V. (1996). Visual function during acute hypoglycemia. *European Journal of Ophthalmology, 6* (1), 81-86.

Taylor, L.A., & Rachman, S.J. (1988). The Effects of Blood Sugar Level Changes on Cognitive Function, Affective State, and Somatic Symptoms. *Journal of Behavioral Medicine, 11* (3), 279-289.

Uitenbroek, D. G. (1997). *SISA Correlations.* Southampton: D.G. Uitenbroek. Retrieved October 3, 2001, from http://home.clara.net/sisa/corrhlp.htm.

Wechsler, D. (1997). *Administration and Scoring Manual for the Wechsler Adult Intelligence Scale - Third Edition.* San Antonio, TX: The Psychological Corporation.

Wechsler, D. (1997). *Technical Manual for the Wechsler Adult Intelligence Scale - Third Edition.* San Antonio, TX: The Psychological Corporation.

Wechsler, D. (1944). *The Measurement of Adult Intelligence* (3rd ed.). Baltimore: Williams & Wilkins.

Appendixes

Appendix A

Aviation Safety Reporting System (ASRS) Search Terms

"pilots and blood sugar"

"pilots and lack of food"

"hungry"

"low blood sugar"

"blood sugar"

"pilot and headache and food"

"pilot and confusion and food"

"blurred vision and pilot"

"pilot and dizziness"

"lack and crew meals"

"pilot and nutrition"

"hypoglycemia"

"pilot and food and hungry"

"crew meals"

"not eaten"

"pilot and crew meal and lack" (redundant files found)

Appendix B

Consent Form

San José State
UNIVERSITY
FOUNDATION

Agreement to Participate in Research - Consent Form

Responsible Investigators: Dr. Kevin Jordan, Jolene M. Bischoff
Title of Protocol: Measurement Design Study

1. I have been asked to participate in a research study investigating measurement design.

2. I will be asked to complete written and verbal measures on campus, at San Jose State University, on ________ (m/dy/yr) between the times of ________ (am-pm) and on ________ (m/dy/yr) between the times of ________ (am-pm).

3. No risks are anticipated while taking part in this study.

4. Participation in this study will contribute to further research in this area.

5. The results of this study may be published, but no information that could identify the participant will be included.

6. Participants will receive 2 hours of credit for taking part in this study.

7. Any questions about the research may be addressed to the principal investigator, Jolene M. Bischoff via e-mail: jbischoff@mail.arc.nasa.gov. Complaints about the research may be presented to the Professor of Psychology Kevin Jordan, (408) 924-5600. Questions about research, subjects' rights, or research-related injury may be presented to Nabil Ibrahim, Ph.D. Associate Vice President for Graduate Studies and Research, at (408) 924-2480.

8. No service of any kind, to which a subject is otherwise entitled, will be lost or jeopardized if a person chooses to "not participate" in the study.

9. Consent is given voluntarily. A participant may refuse to take part in this study or in any part of this study. If a participant decides to take part in this study, he or she is free to withdraw at any time without prejudice to the participant's relations with San Jose State University or any other participation institutions.

10. The participant has received a signed and dated copy of this consent form.

The signature of a participant on this document indicates agreement to be included in the study. The signature of a researcher on this document indicates agreement to include the above named participant in the research and attestation that the participant has been fully informed of his or her rights.

Signature ____________________ Date

Investigator's Signature ____________________ Date

Appendix C

Demographics Sheet

I.D. # ______________________________

Age: ________ years

Please circle your responses.

Gender: Male Female

Level of Education (class level):

Freshman

Sophomore

Junior

Senior

Ethnicity:

African-American	Asian-American	Latino
Caucasian	Native American	Other ______________

Is English your first language?	Yes	No
Do you wear glasses to correct your vision?	Yes	No
Do you have a history of brain injury?	Yes	No
Do you have any learning disabilities?	Yes	No

Are you right-handed, left-handed, or ambidextrous?

Right-handed Left-handed Ambidextrous

Appendix D

Raw Score Conversion & Summary Sheets

Participant ID:		
Gender:	Male	Female
Age: (years)		

Digit Symbol Coding	**Raw Scores**	**DSC Scaled Scores**
Version 1 (original)		
Version 2		
Version 3		
Version 4		

Letter-Number Sequencing	**Raw Scores**	**LNS Scaled Scores**
Version 1 (original)		
Version 2		
Version 3		
Version 4		

Trail Making B	**Total Time**	**# of Errors**
Version (original)		
Version 1		
Version 2		
Version 3		
Version 4		

Signed Approval Form

San José State
UNIVERSITY

Office of the Academic Vice President

Associate Vice President Graduate Studies and Research

One Washington Square
San José, CA 95192-0025
Voice: 408-283-7500
Fax: 408-924-2477
E-mail: gstudies@wahoo.sjsu.edu
http://www.sjsu.edu

To: Jolene M. Bischoff
4000 Rio Road #15
Carmel, CA 93923

From: Nabil Ibrahim
AVP, Graduate Studies & Research

Date: March 18, 2002

The Human Subjects-Institutional Review Board has granted a request to change the title for an approved research study. The protocol was originally reviewed under the title, "The Glucose Cognitive Assessment Tool (G-CAT): Establishment of Alternate Forms," and was approved on January 25, 2002. The study is now entitled:

"Measurement Design: Developing Reliable Alternate Forms of Cognitive Tests."

Since there are no major changes to the protocol, the Human Subjects-Institutional Review Board has granted this project a one-year extension effective from the date of the original approval. Data collection beyond January 24, 2003 requires an extension request.

If you have any questions, please do not hesitate to contact me at (408) 924-2480.