

AFRL-SA-BR-TR-2010-0002



**Joint use of the MAB – II and MicroCog for
Improvements in the Clinical and Neuropsychological
Screening and Aeromedical Waiver Process of
Rated USAF Pilots**



**Wayne Chappelle, Psy. D., ABPP
United States Air Force School of Aerospace Medicine
Aerospace Medicine Consultation Division
Aerospace Clinical Psychology Function
Brooks City-Base, Texas**

**Malcolm J. Ree, Ph.D.
Erica L. Barto, M.A.
Mark S. Teachout, Ph.D.
Operational Technologies Corporation
San Antonio, Texas**

**William T. Thompson, M.A.
Eagle Applied Sciences
San Antonio, Texas**

January 2010

Final Report

**Approved for public release, distribution
unlimited. Approved by 311th/ABG Public
Affairs Office, Case file no. 10-061, 25 Feb
2010, Brooks City-Base, Texas 78235.**

**Air Force Research Laboratory
711th Human Performance Wing
Aerospace Medicine Department
Aerospace Medicine, Consultation Division
2507 Kennedy Circle
Brooks City-Base, Texas 78235-5116**

NOTICES

When Government drawings, specifications, or other data are used for any purpose other than in connection with a definitely Government-related procurement, the United States Government incurs no responsibility or any obligation whatsoever. The fact that the Government may have formulated or in any way supplied the said drawings, specification, or other data, is not to be regarded by implication, or otherwise in any manner construed, as licensing the holder or any other person or corporation; or as conveying any rights or permission to manufacture, use, or sell any patented invention that may in any way be related thereto.

The mention of trade names or commercial products in this publication is for illustration purposes and does not constitute endorsement or recommendation for use by the United States Air Force.

The Office of Public Affairs has reviewed this report, and it is releasable to the National Technical Information Service, where it will be available to the general public, including foreign nationals.

This report has been reviewed and is approved for publication.

Government agencies and their contractors registered with Defense Technical Information Center (DTIC) should direct requests for copies to: Defense Technical Information Center, 8725 John J. Kingman Rd., STE 0944, Ft. Belvoir, VA 22060-6218.

Non-government agencies may purchase copies of this report from: National Technical Information Services (NTIS), 5285 Port Royal Road, Springfield, VA 22161-2103

 //Signed//

WAYNE CHAPPELLE, Psy.D., ABPP
Aerospace Clinical Psychology Function

 //Signed//

KEITH E. BRANDT, Col. USAF, MC, CFS
Chief, Aerospace Medicine Consultation Division

REPORT DOCUMENTATION PAGE			<i>Form Approved</i> <i>OMB No. 0704-0188</i>	
Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing this collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number. PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.				
1. REPORT DATE (DD-MM-YYYY) 01-01-2010		2. REPORT TYPE FINAL TECHNICAL	3. DATES COVERED (From – To) N/A	
4. TITLE AND SUBTITLE Joint Use of the MAB-II and MicroCog for Improvement in the Clinical and Neuropsychological Screening & Aeromedical Waiver Process of Rated USAF Pilots			5a. CONTRACT NUMBER	
			5b. GRANT NUMBER	
			5c. PROGRAM ELEMENT NUMBER	
6. AUTHOR(S) Wayne L. Chappelle, Malcolm J. Ree*, Erica L. Barto*, Mark S. Teachout*, and William T. Thompson**			5d. PROJECT NUMBER	
			5e. TASK NUMBER	
			5f. WORK UNIT NUMBER	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Air Force Materiel Command, Air Force Research Laboratory, 711 Human Performance Wing, Aerospace Medicine Department, Aerospace Medicine Consultation Division, 2507 Kennedy Circle, Brooks City-Base, Texas 78235-5116 *Operational Technologies Corporation, 4100 NW Loop 410, San Antonio, Texas 78229-4254 **Eagle Applied Sciences, Inc. 200 Concord Plaza, Suite 430, San Antonio, Texas 78229-4254			8. PERFORMING ORGANIZATION REPORT NUMBER AFRL-SA-BR-TR-2010-0002	
			10. SPONSOR/MONITOR'S ACRONYM(S)	
			11. SPONSOR/MONITOR'S REPORT NUMBER(S)	
9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)				
12. DISTRIBUTION / AVAILABILITY STATEMENT Approved for public release, distribution unlimited, Approved by 311 th /ABG Public Affairs Office, case file no. 10-061, 25 Feb 10, Brooks City-Base, Texas 78235				
13. SUPPLEMENTARY NOTES				
14. ABSTRACT Intelligence testing and neuropsychological screening is critical to USAF medical flight screening and aeromedical waiver procedures for pilots. A confirmatory factor analysis was conducted on the intelligence test scores from the Multidimensional Aptitude Battery (MAB-II) and neuropsychological screening from the five first-level indices of the MicroCog. The sample consisted of 10,612 USAF pilot applicants selected for training. Subtest correlations and measurement models were estimated for both tests. A joint factor analysis using the measurement models of the MAB-II and MicroCog showed positive correlations between factors identified for the two tests and between subtests measuring specific cognitive aptitudes. Similarities and differences among scores are discussed. Normative tables demonstrate cognitive aptitudes for USAF pilot applicants are in the high average to superior range when compared to the general population. The normative tables improve interpretation of scores for USAF pilots and an aeromedical clinical case vignette is provided to illustrate practical applications of the study.				
15. SUBJECT TERMS Intelligence testing, neuropsychological screening, USAF pilot, medical flight screening, aeromedical evaluations				
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT SAR	18. NUMBER OF PAGES 37
a. REPORT U	b. ABSTRACT U	c. THIS PAGE U		
			19b. TELEPHONE NUMBER 210-536-3143	

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std. Z39.18

THIS PAGE LEFT INTENTIONALLY BLANK

TABLE OF CONTENTS

INTRODUCTION	1
Cognitive Assessment of USAF Pilots.....	1
Medical Flight Screening – Cognitive Baseline Testing	2
Aeromedical Utilization.....	3
Purposes of the Study	3
METHOD.....	5
Participants.....	5
Measures.....	5
<i>MAB II</i>	5
<i>MicroCog</i>	7
Procedure	8
RESULTS	10
Descriptive Statistics	10
Comparison of sample data to normative values	10
Comparison of sample data to previous research	11
Normative Distribution of Scores for Air Force Pilots	11
Confirmatory Factor Analysis Results.....	13
DISCUSSION.....	18
Subtest Correlations within the <i>MAB II</i> and <i>MicroCog</i>	18
<i>MAB - II</i> subtest correlations.....	18
<i>MicroCog</i> first – level index correlations.....	19
Subtest Correlations Between the <i>MAB – II</i> and <i>MicroCog</i>	19
Pattern Analyses Between <i>MicroCog</i> and <i>MAB – II</i> Subtests.....	20
Confirmatory Factor Analysis	21
Aeromedical Case Study Application.....	21

LIMITATIONS TO THE 'BEST FIT MODEL'25
CONCLUSION26
REFERENCES27

LIST OF TABLES

TABLE 1. <i>MAB II</i> Factors, subtests and descriptions	6
TABLE 2. <i>MicroCog</i> Index descriptions.....	7
TABLE 3. Comparison of <i>MAB – II</i> and <i>MicroCog</i> Normative and Current Sample scores).....	10
TABLE 4. Distribution of T – Scores from the <i>MAB – II</i> for Air Force Pilots (N = 10,612).....	12
TABLE 5. Distribution of Age and Education Corrected Normative Standardized Scores from the <i>MicroCog</i> for the Sample of Air Force Pilots.....	12
TABLE 6. Correlations of the all the Variables (<i>MAB – II</i> Subtests and <i>MicroCog</i> First – Level Indices)...	13
TABLE 7. Goodness-of-Fit Statistics for the Two Measurement Models.....	14
TABLE 8. Case Example: Baseline and Post-injury Performance on the <i>MAB – II</i> and <i>MicroCog</i>	23

LIST OF FIGURES

Figure 1. Confirmatory Factor Analysis of the *MAB – II*.....15

Figure 2. Factor Structure of the *MicroCog*.....16

Figure 3. Factor Structure of *MicroCog* with Two Factors.....16

Figure 4. Joint Factor Structures of the *MAB – II* and *MicroCog*.....17

Joint use of the MAB – II and MicroCog for Improvements in the Clinical and Neuropsychological Screening and Aeromedical Waiver Process of Rated USAF Pilots

INTRODUCTION

USAF pilots are in a high-risk profession where mistakes in the performance of their flying duties can lead to significant costs in terms of human life, national security, and foreign relations. Furthermore, each year there are several aviation related mishaps in training and operational missions where human error is often considered a causal factor. As a result of the high risk nature of military flying and the evidence that a pilot's cognitive disposition is crucial for safe and effective flying, the cognitive assessment of USAF pilots is a critical part of medical flight screening and the aeromedical waiver process (Olea & Ree, 1994; Ree & Carretta, 1996; Ree, Carretta, & Teachout, 1995).

Currently, the assessment of cognitive functioning prior to and following pilot training is composed of two tests: the *Multidimensional Aptitude Battery - Second Edition (MAB – II)* and the *MicroCog*. These tests are integral to the comprehensive clinical and neuropsychological evaluations at the Aeromedical Consultation Service (ACS) of the United States Air Force School of Aerospace Medicine (USAFSAM). The results of these tests are especially important to pilots and pilot candidates seeking a waiver for flying due to a history of an illness or injury that may have affected their cognitive disposition. It is critical to clinical and neuropsychological aeromedical evaluations to have a clear understanding of the relationship between the two tests. Such an understanding is essential to the interpretation and pattern analyses of subtest and index scores when assessing intellectual performance and screening for general or specific cognitive deficits. However, there are no published studies that have assessed the relationships between these two tests.

Cognitive Assessment of USAF Pilots

Prior to entering pilot training, each candidate completes psychological testing to obtain a baseline of their cognitive disposition. The purpose of this testing is to have an accurate assessment of a pilot's general cognitive functioning, as well as specific cognitive aptitudes. This baseline evaluation is made just prior to pilot training and typically during medical flight screening. The scores from baseline testing are held in reserve to serve as a basis for comparison should the pilot need to undergo a psychological evaluation at a later date during training or afterwards. If a pilot has a history of suffering from a medical condition (e.g., bacterial meningitis, closed head injury, multiple sclerosis) or a psychiatric illness (e.g., major depression, anxiety) that affected his or her psychological disposition, he or she must receive an aeromedical psychological evaluation to determine how well they have recovered. Further, the

aeromedical psychological evaluation must determine if there are any changes in a pilot's or pilot candidate's cognitive disposition that would raise concerns if he or she were to return to flying. Without baseline testing, it is very difficult to make such determinations. Aeromedical evaluations are highly sensitive, and even subtle changes in a pilot's cognitive functioning can cause disqualification from flying. Therefore, having an accurate baseline of a pilot's cognitive functioning is critical to a pilot's future, the aeromedical waiver process, and mission readiness.

Medical Flight Screening- Cognitive Baseline Testing

At the present time, baseline testing of a pilot's cognitive disposition is composed of two computerized tests, the *MAB – II* and the *MicroCog*. The *Multidimensional Aptitude Battery (MAB)*, developed in 1984, was designed to measure general cognitive ability through five performance and five verbal scores. These scores were then combined to yield a full scale IQ score (Jackson, 1984). Previous research has demonstrated that this full scale IQ score measures general mental ability in a multitude of age groups (Wallbrown, Carmin, & Barnett, 1988, 1989; Lee, Wallbrown, & Blaha, 1990; Kranzler, 1991; Carretta, Retzlaff, Callister, & King, 1998).

Specifically, Retzlaff and Gibertini (1988) administered the *MAB* to a sample (n = 350) of USAF student pilots. They found that the full scale, verbal, and performance IQs were approximately 1 to 1.3 standard deviations above the population mean. Retzlaff, Callister, and King (1999) found similar results in a sample of 537 USAF student pilots a decade later.

The *MAB* was reviewed and re-standardized in 2003 to ensure that it continued to be an effective measure of general cognitive ability. The result was the *Multidimensional Aptitude Battery – II (MAB – II)*, a version of the test that could be scored, interpreted, and administered to an individual or group of individuals via computer (Jackson, 2003). Retzlaff, King, and Callister (1995) compared a paper-and-pencil version of the *MAB* to the computerized version and did not find a significant difference between the two tests. In the USAF, the *MAB – II* has been used to assist clinicians in evaluating pilots and pilot applicants.

In addition to its clinical application, professional aviation psychology practice indicates that *MAB – II* scores (Callister, King, & Retzlaff, 1996) might be useful in detecting applicants likely to fail pilot training (King & Flynn, 1995) despite qualifying through USAFA, OTS, or ROTC. Carretta, Retzlaff, and King (1997), examined the extent to which the Air Force Officer Qualifying Test (AFOQT), a cognitive test currently used by the USAF (Carretta & Ree, 1996) and the *MAB*, measure the same constructs. A joint factor analysis revealed that both batteries had hierarchical structure. The higher order factor in the AFOQT has been identified previously as general cognitive ability (*g*). The intercorrelations between the higher order factors from the two batteries was .981, demonstrating that both measured *g*. Since there is considerable overlap between these two batteries, AFOQT and *MAB – II*, similar research is needed to determine the overlap between the *MAB – II* and the *MicroCog*.

The *MicroCog* is administered and scored via computer and assesses key neurocognitive functioning in adults (Powell, Kaplan, Whitla, Weintraub, Caitlin, & Funkenstein, 1993). It was designed as a screening device and a diagnostic tool to be used during a neuropsychological or clinical examination for the assessment of cognitive functioning. The four primary goals of *MicroCog* development were to: (a) base the test on a model of brain-behavior relationships, (b) measure information processing speed, (c) attend to the interaction between education and performance, and (d) extend norms into the ninth decade of life. The *MicroCog*, developed in 1993, was designed to measure the five content-specific domains of attention/mental control, memory, reasoning/calculation, spatial processing, and reaction time. Scores for these five domains are the sum of two to seven specific subtests.

Aeromedical Utilization

Both the *MAB – II* and *MicroCog* are administered at the Aeromedical Consultation Service's (ACS) Department of Neuropsychiatry. Such tests are part of a comprehensive clinical psychological and/or neuropsychological assessment of USAF pilots and, at times, pilot candidates. USAF pilots or pilot candidates are referred to the Department of Neuropsychiatry for such evaluations when they have been disqualified. As mentioned previously, disqualification is often due to a medical or psychiatric illnesses that potentially affect cognitive disposition. Performance on the *MAB – II* and *MicroCog* during an ACS evaluation is compared with baseline scores, collected prior to pilot training to assess for changes that have occurred over time and are due to a history or presence of an aeromedically disqualifying condition. Impairments in cognitive functioning may be reflected in responses to specific subtests, indices, pattern analyses, and differences between current and baseline test scores.

Purposes of the Study

Based on the review of the literature, this study investigated relationships between the *MAB – II* and the *MicroCog*. The first purpose was to compare the pilot trainee scores in this sample for both the *MAB – II* and the *MicroCog* to the nationally representative normative scores for each test. The second purpose was to compare descriptive statistics (e.g., mean scores and standard deviations) of the current sample to those obtained in previous research. The third purpose was to examine the similarity of the constructs measured by the two tests.

Although the descriptions of the *MAB – II* and the *MicroCog* suggest similarities and differences, empirical evidence of the nature of these similarities and differences has not been investigated, particularly in highly selected samples, such as pilots. Since these relationships have seldom been examined in the literature, doing so is crucial to accurate aeromedical evaluations. Although there are specific descriptions of each subtest and index scores, the

degree of statistical correlation among the subtests and indices between the two tests is unknown. Because of the importance of these tests in aeromedical clinical and neuropsychological evaluations of pilots, the sensitive nature of the aeromedical waiver process, and the implications for mission readiness and safety, it is essential to have an accurate and clear understanding of the statistical relationships between scores on these tests. This will significantly improve the interpretation of the pattern analyses of scores within and between the *MAB – II* and *MicroCog*. For example, do scores on the spatial subtest of the *MAB – II* measure a construct similar or identical to the spatial subtest on the *MicroCog*? Are similarly described subtests interchangeable? This study will help us understand the factorial relationship between the two tests and will provide current descriptive normative data based upon USAF pilots and pilot candidates for both tests. This information will enhance the clinical and neuropsychological evaluations of those in need of an aeromedical waiver.

METHOD

Participants

A total of 10,612 pilot candidates going through medical flight screening were included in this study. In the sample, 91% were male and 9% were female. Participants had a mean age of 22 years (SD= 2.7). Furthermore, 84% were Caucasian, 4% were Hispanic, 2% were African American, 6% were “other,” and 4% “not reporting.” All were either college graduates or enrolled in their 4th year of college. All participants were found to be physically and psychologically healthy and to have met the enhanced aeromedical standards required for attending pilot training and becoming a rated USAF pilot.

Measures

MAB-II. The *MAB – II* is a broad-based test of cognitive functioning (*MAB – II*; Jackson, 1998, 2003). The content and structure of the test was fashioned after the *Wechsler Adult Intelligence Scale (WAIS-III*; Wechsler, 1997), which is the most widely used individually administered test of cognitive functioning and intelligence. The *MAB – II* has 10 subtests that are each seven minutes long and all items have five multiple-choice responses. The *MAB – II* requires only 70 minutes to complete and can be administered in group settings. Administration of this test produces verbal (VIQ), performance (PIQ) and full scale (FSIQ) intelligence quotient scores, which are global measures of cognitive functioning. The test is separated into Verbal abilities (i.e., subtests of information, comprehension, arithmetic, similarities, and vocabulary) and Visual-Performance based abilities (i.e., subtests of digit symbol coding, picture completion, spatial analyses, picture arrangement, and object assembly). The *MAB – II* normative subtest scores for the general population have a mean of 50 and a standard deviation of 10. The normative VIQ, PIQ, and FSIQ scores in the general population have a mean of 100 and a standard deviation of 15. The *MAB – II* manual has well documented internal consistency, test-retest reliability, and validity coefficients. Table 1 contains a description of the factors measured, subtests associated with each factor, and the abbreviations for the factors and subtests.

TABLE 1. *MAB-II* Factors, Subtests, and Descriptions.

Factor	Subtests	Description
Verbal Intelligence Quotient (FVIQ)	Information (inf)	General fund of knowledge; long-term memory
	Comprehension (com)	Social reasoning and comprehension
	Arithmetic (ari)	General and numerical reasoning; problem solving
	Similarities (sim)	General reasoning and problem solving
	Vocabulary (voc)	Flexibility and adjustment to novelty, reasoning, abstract thought, long-term memory
Performance Intelligence Quotient (FPIQ)	Digital Symbol (ds)	Adaptation to new set of demands; visual learning and coding, figural memory, and speed of information processing
	Picture Completion (pc)	Visual attention to detail; knowledge of common objects; perceptual and analytical skills
	Spatial Score (sp)	Ability to visually and mentally rotate abstract two-dimensional images of objects in different positions; figural-domain reasoning
	Picture Arrangement (pa)	Visual reasoning; ability to identify a meaningful sequence; social intelligence; perceptual reasoning
	Object Assembly (op)	Visualization and visuo-construction skills; perceptual analytical skills needed to identify a meaningful object from left-to-right sequence

MicroCog. The second computer-based test is the *MicroCog* (Powell et al., 1993). This test was designed to screen specific neurocognitive aptitudes. It is composed of 18 subtests that are, in turn, combined to yield five first-level indices (i.e., attention and mental control, reasoning and executive functioning, memory, spatial analyses, and reaction time). The scores from the five first-level indices are aggregated to compute second-level indices regarding speed of information processing (SIP) and accuracy of information processing (AIP). The scores from the second-level indices are then re-weighted to create third-level indices for general cognitive functioning and general cognitive proficiency (i.e., a weighted assessment of a person’s overall cognitive functioning that also accounts for speed of information processing). The second and third levels are differentially weighted aggregates of scores from the first-level indices. The *MicroCog* is individually administered and scored via a computer. Multiple choice items with varying numbers of response options and free-response items are presented for each test. General speed of information processing is measured by reaction time to each item. The five first-level indices, two second-level indices and two third-level indices of the *MicroCog* each have a mean of 100 and standard deviation of 15. These scores have been statistically adjusted for age and level of education. The manual for the *MicroCog* has well documented internal consistency, test-retest reliability, and validity coefficients (Christal, 1990; Kyllonen, 1985, 1991; Powell et al., 1993). Table 2 contains a description of the general neuropsychological aptitudes and their abbreviations for the *MicroCog* first-level indexes.

TABLE 2. *MicroCog* Index Descriptions.

Index	Description
Spatial Processing (spatial)	Memory for novel spatial arrangements, visuo-perceptual ability
Attention/Mental Control (atten)	Concentration, span of attention, diligence, persistence, resistance to interference
Reasoning/Calculation (reason)	Inductive reasoning, cognitive flexibility, concept formation, basic arithmetic
Memory (memory)	Short-term memory (storing information for a brief period) and long-term memory (storing information for a longer time period, from minutes to years)
Reaction Time (react)	Length of psychomotor time between presented stimulus and response, readiness to respond, vigilance, attention

Procedure

The sample of pilot candidates in this study were administered both the *MAB – II* and the *MicroCog* as a routine part of medical flight screening prior to attending pilot training. The variables chosen for analysis were those that are reported and interpreted in clinical and other settings. These included the ten subtests and three composites of the *MAB – II* and the five first-level index scores of the *MicroCog*. Furthermore, level-two and level-three index scores of the *MicroCog* cannot be analyzed simultaneously with the first-level index scores. This is because the second- and third-level index scores are differentially weighted composites of the first-level index scores and, as such, a matrix including all nine scores (five first-level indexes, two second-level indexes, and two third-level indexes) is not amenable to statistical techniques essential to the factor analysis conducted¹.

Descriptive statistics were computed for the *MAB – II* subtest and IQ scores, as well as the *MicroCog* first-level index scores. These included means and standard deviations. Additionally, the data were examined for missing, out of range, or inappropriate values. Participants with erroneous data were removed. Twelve subjects were removed for this reason. Furthermore, distribution of subtest T-Scores from the *MAB – II* and standardized first-level index scores from the *MicroCog* were developed to assess how the distribution of scores differed from the civilian, non-aircrew general population. See Tables 3 and 4.

Confirmatory factor analyses were conducted in three interdependent steps using AMOS 17 (Arbuckle, 2007; Blunch, 2008). Each estimation was accomplished using covariance matrices and maximum likelihood methods. First, a confirmatory factor analysis (CFA) measurement model was specified and evaluated for the ten subtests of the *MAB – II*. Step two specified and tested two measurement models of the scores from the *MicroCog*. The third and final step was to apply the two measurement models (*MAB – II* and *MicroCog*) to the sample ($n = 10,612$) which had taken both test batteries to determine the relationships among the factors.

In testing the two measurement models, the following goodness-of-fit indices were used: comparative fit index (CFI), goodness-of-fit index (GFI), root mean square error of approximation (RMSEA), and critical N (CN). Based on best practices (Lance & Vandenberg, 2009), the following values were established as minima for an acceptable fit: $CFI \geq .90$, $GFI \geq .90$, $RMSEA \leq .08$, and $CN \geq 200$. The model χ^2 was also estimated, but owing to its great statistical power, was not evaluated. It is presented because many of the other goodness-of-fit indices are based on χ^2 .

¹ Technically, the covariance matrix must be invertible. The second- and third-level indices are redundant with the first-level indices and yield a non-positive-definite Hessian matrix, prohibiting inversion of the covariance matrix (Neudecker & Magnus, 1988).

The measurement model of the *MAB – II*, shown in Figure 1, was suggested by the test developer and called for two factors corresponding to the Verbal IQ and Performance IQ. Table 1 shows which subtests comprise the verbal factor (FVIQ) and the performance factor (FPIQ). The letter “F” in the abbreviations indicates a factor, not a summed composite of scores.

Two models of the *MicroCog* factors were fit and tested against each other. The first model (shown in Figure 2) was a one-factor solution. The second model (shown in Figure 3) was a two-factor solution suggested by a description of the content and process similarity of the level-one indexes, “spatial” and “attention.”

Both *MicroCog* factor models were estimated. To test the difference between the models, χ^2 was estimated for each model. The difference between these two χ^2 values was divided by the difference in the degrees of freedom between the two models and the result evaluated at $p < .05$. A significant χ^2 would indicate a difference in the two models and require the use of the more complex model. A non-significant χ^2 would allow the less complex model to be used.

The correlations of all the factors were estimated in the sample of participants who had taken both test batteries (N = 10,612). Because the measurement models were evaluated independently with the goodness-of-fit measures, no estimates of goodness-of-fit were informative for the model having both factor structures in which correlations were estimated.

To estimate the proportion of general cognitive ability (*g*) measured by each test, *MAB – II* and *MicroCog*, the first unrotated principal factor was computed. This was done independently for the *MAB – II* and independently for the *MicroCog* (Ree & Earles, 1991). The percentage of variance associated with the first unrotated principal factor is an indication of how much the test measures general cognitive ability. The joint factor structure is shown in Figure 4.

RESULTS

Two sets of analyses were conducted to address the purposes of the study. The first set of analyses calculated descriptive statistics that were used to compare the pilot scores in this sample for both the *MAB – II* and the *MicroCog* to the nationally representative normative scores. The descriptive statistics for the current sample were then compared to those obtained in previous research (Retzlaff, Callister, & King, 1999; Retzlaff & Gibertini, 1988). The second set of analyses examined the similarity of the constructs measured by the two tests.

Descriptive Statistics

Means and standard deviations calculated for the *MAB – II* and *MicroCog* are contained in Table 3. Table 3 also contains percentages of the normative values for each IQ composite of the *MAB – II* and each first-level index of the *MicroCog*, permitting comparisons of the present sample to these norms. These sample data can also be compared to sample data from previous studies.

TABLE 3. Comparison of *MAB – II* and *MicroCog* Normative and Current Sample Scores

	Mean Score	% of Normative	S.D.	% of Normative
Scores of the <i>MAB – II</i>				
Verbal IQ	119.42	119	6.66	44
Performance IQ	119.90	120	8.05	54
Full Scale IQ	121.02	121	6.46	43
Indexes of the <i>MicroCog</i>				
Attention/Mental Control	103.04	103	12.87	86
Memory	110.62	110	13.57	90
Spatial Processing	106.67	107	10.29	68
Reasoning/Calculation	97.26	97	12.97	86
Reaction Time	97.70	98	12.65	84

Comparisons of sample data to normative values. The first purpose of this study was to compare the pilot trainee scores in this sample for both the *MAB – II* and the *MicroCog* to the nationally representative normative scores for each test. For the *MAB – II*, the Full Scale, Verbal, and Performance IQ scores in the nationally representative normative sample are standardized, with means and standard deviations of 100 and 15, respectively. The data in Table 1 show that the means of the current sample are elevated, ranging from 119% to 121% of the normative sample. The standard deviations for the current sample are reduced, ranging from 43% to 54% of the normative sample.

For the *MicroCog*, the five first-level indices are reported, because these are used operationally. These scores are also based on a nationally representative normative sample, but unlike the *MAB – II*, have been statistically adjusted for age and educational level. Each of these indexes is then standardized with a mean of 100 and standard deviation of 15. The data in Table 3 show that the means were elevated for three of the indexes, Attention/Mental Control, Memory, and Spatial Processing, ranging from 103% to 111% of the normative-sample value, while the means were reduced for two of the indexes. The means for Reasoning/Calculation and Reaction Time were 87% and 98% of the normative-sample value, respectively. Furthermore, each of the five standard deviations from the current sample was reduced, ranging from 68% to 90% of the normative-sample value. The reduction in the standard deviations in both tests is a consequence of a highly selected sample and is called range restriction.

Comparison of sample data to previous research. The second purpose of this study was to compare descriptive statistics (e.g., mean scores and standard deviations) of the current sample to those obtained in previous research. Results indicate that the means and standard deviations from the current sample are very similar to findings by Retzlaff, Callister, and King, (1999) who found a mean full scale IQ score of 119. Retzlaff and Gibertini (1988) found a mean full scale IQ score of 120. USAF student pilots are a bright sample with scores ranging from 1.0 to 1.3 standard deviations above the population mean. Unfortunately, there is no previous published research using a pilot sample for the *MicroCog*, so there is no basis for comparing the current results with previous results.

Normative Distribution of Scores for Air Force Pilots

The normative distributions of scores for Air Force pilots for the *MAB – II* and *MicroCog* are contained in Tables 4 and 5, respectively. Table 4 provides the percentile equivalent scores according to the T-scores for the subtests of the *MAB – II*. T-scores have mean of 50 and a standard deviation of 10. Therefore, a T-score of 60 is one standard deviation above the mean and a T-score of 30 is two standard deviations below the mean. If the sample of student pilots was just equivalent to the normative population, their scores in the column labeled 50th would have entries of 50. However, the smallest entry is 51 and the largest entry is 69 indicating that this sample clearly scored higher than the normative population. The table shows the T-scores at specific percentiles. For example, in the first row, a T-score of 56 on the *MAB – II* information subtest corresponds with the 5th percentile based on the distribution of scores for Air Force pilots.

TABLE 4. Distribution of T-Scores from the *MAB – II* for Air Force Pilots (N = 10,612).

Subtests	<i>Percentile</i>								
	MIN	5 th	10 th	25 th	50 th	75 th	90 th	95 th	MAX
Information	39	56	59	63	68	71	74	75	83
Comprehension	29	51	55	57	61	63	64	66	72
Arithmetic	33	49	52	58	61	66	70	74	89
Similarities	37	53	54	57	61	64	65	67	74
Vocabulary	38	49	51	55	60	64	68	71	85
Digit Symbol	24	55	59	63	69	71	74	74	76
Picture Completion	23	48	52	55	61	65	67	69	75
Spatial	21	50	52	57	61	65	69	71	75
Picture Arrangement	17	41	41	47	51	58	61	64	81
Object Assembly	30	53	55	59	63	65	67	68	71

Table 5 shows the percentile equivalents for the first-level index scores for the *MicroCog*. For example, in the first row, Attention/Mental Control, an index score of 122 corresponds to the 95th percentile based on the distribution of scores for Air Force pilots. The scores have a mean of 100 and a standard deviation of 15. Therefore, a score of 115 is one standard deviation above the mean and a score of 70 is two standard deviations below the mean. If the sample of student pilots was just equivalent to the normative population, their scores in the column labeled 50th would all be 100. The smallest entry is 98 and the largest entry is 112. Overall, the Air Force pilots were a high scoring sample compared to the normative population.

TABLE 5. Distribution of Age and Education Corrected Normative Standardized Scores from the *MicroCog* for the Sample of Air Force Pilots (N = 10,612).

First-Level Indexes	<i>Percentile</i>								
	MIN	5 th	10 th	25 th	50 th	75 th	90 th	95 th	MAX
Attention/Mental Control	50	79	87	96	104	111	117	122	150
Reasoning/Calculation	50	75	80	89	98	107	113	116	139
Memory	51	86	92	101	112	121	127	130	143
Spatial Processing	54	86	93	102	109	114	117	119	135
Reaction Time	50	74	81	91	100	106	112	115	147

Confirmatory Factor Analysis Results

The third purpose of this study was to examine the similarity of the constructs measured by the two tests, *MAB – II* and the *MicroCog*. Parameter estimates for the Confirmatory Factor Analysis (CFA) converged quickly, indicating no statistical problems with the data. Table 6 shows the correlations of all the variables for the sample of 10,612 participants who were administered both the *MAB – II* and the *MicroCog*.

TABLE 6. Correlations of all the Variables (*MAB – II* Subtests and *MicroCog* First-Level Indices).

	Inf	com	ari	sim	voc	ds	pc	sp	pa	op	spatial	atten	reason	memory	react
inf	1.00														
com	.43*	1.00													
ari	.26*	.28*	1.00												
sim	.48*	.41*	.25*	1.00											
voc	.58*	.44*	.27*	.53*	1.00										
ds	.12	.12	.27*	.18	.11	1.00									
pc	.28*	.22	.12	.24*	.21	.22	1.00								
sp	.14	.12	.28*	.14	.09	.34*	.29*	1.00							
pa	.18	.14	.14	.19	.15	.25*	.29*	.27*	1.00						
op	.22	.18	.23	.21	.18	.33*	.41*	.40*	.33*	1.00					
spatial	.07	.09	.17	.09	.05	.26*	.17	.23	.14	.20	1.00				
atten	.16	.19	.26*	.20	.18	.31*	.14	.21	.13	.19	.30*	1.00			
reason	.23	.19	.37*	.23	.19	.32*	.22	.30*	.24*	.28*	.24*	.31*	1.00		
memory	.29*	.26*	.27*	.29*	.30*	.27*	.15	.15	.16	.16	.21	.29*	.31*	1.00	
react	.02	.05	.09	.06	-.01	.20	.18	.16	.18	.15	.17	.15	.17	.08	1.00

Note: * Indicates moderate (.24 - .36) and large (.37 and above) correlations. N = 10,612.

The correlations in Table 6 are positive (except for one, $r = -.01$, n.s.) showing the test scores were all related to one another in the expected fashion. This matrix is an example of “positive manifold” frequently found in ability tests.

The results for the *MAB – II* were straightforward. The model fit was suggested by the publisher and consisted of the two correlated factors interpreted as Verbal and Performance. The following subtests were on the Verbal factor: Information, Comprehension, Arithmetic, Similarities, and Vocabulary. The other five, Digit Symbol, Picture Completion, Spatial, Picture Arrangement, and Object Assembly, were on the Performance factor. As shown in Table 7, the fit of the model to the *MAB – II* data was acceptable. Table 7 shows the fit statistics for the models described in Figures 1, 2, and 3.

TABLE 7. Goodness-of-Fit Statistics for the Two Measurement Models.

Statistic	<i>MAB - II</i>	<i>MicroCog</i> One-Factor	<i>MicroCog</i> Two-Factor
χ^2 (df)	1,594.507 (34)	3.509 (2)	105.600 (4)
GFI	0.97	1.00	0.99
CFI	0.93	1.00	0.98
RMSEA	0.066 (0.063-0.069)	0.008 (0.000 - 0.023)	0.049 (0.041 - 0.057)
CN	324 ($p < .05$)	18,118 ($p < .05$)	954 ($p < .05$)

To determine whether the two-factor model was superior to the one-factor model for the *MicroCog*, χ^2 analyses were conducted (Blunch, 2008). As shown in the first row of Table 7, the χ^2 for the one-factor model was 3.509 with 2 degrees of freedom and the χ^2 for the two-factor model was 105.600 with 4 degrees of freedom. The difference of 102.091 (105.600 – 3.509) with 2 (4-2) degrees of freedom was tested at $p < .05$ and found to be statistically significant. Therefore, the one-factor model, with the superior fit, was preferable and used in the final step of the analyses.

Figure 1 shows the model of the *MAB – II*, factor loadings, and factor correlation.

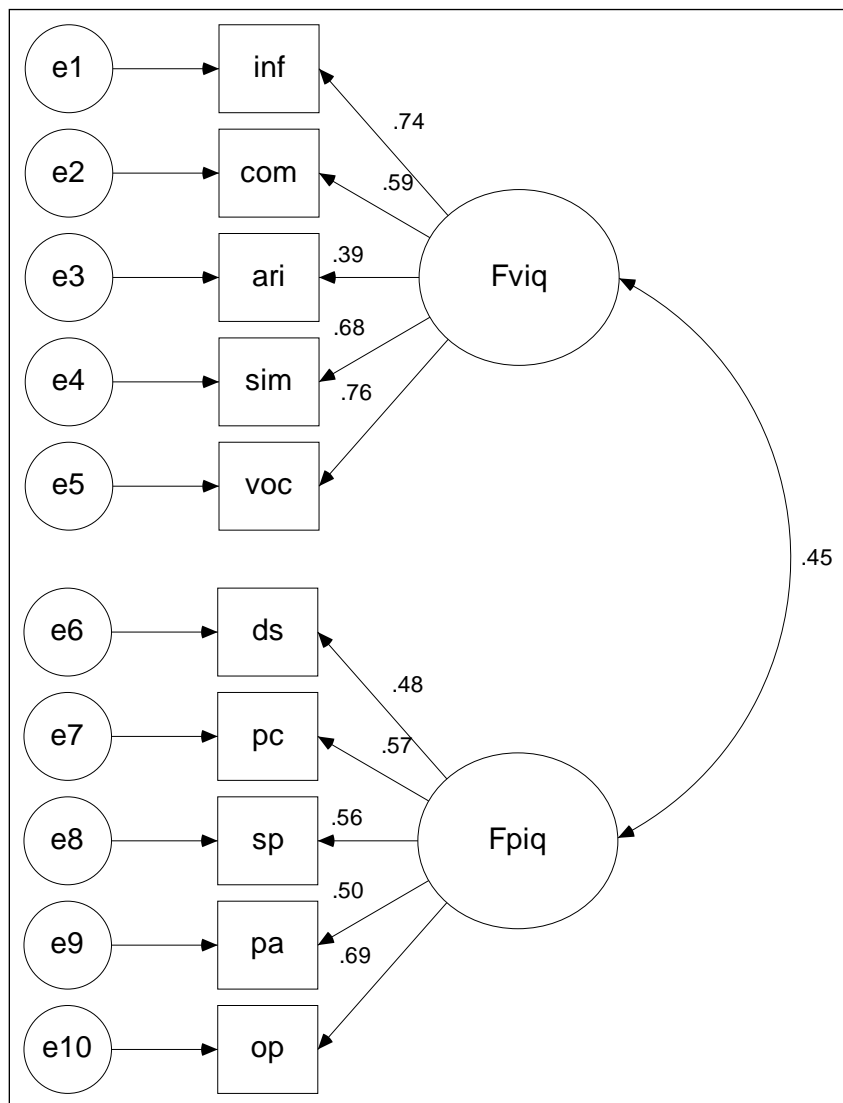


Figure 1 Confirmatory Factor Analysis of the *MAB – II*

The results of the CFA for the *MicroCog* were provided for two models: one-factor and two-factors. Figure 2 shows the one-factor model of the *MicroCog* and factor loadings. Figure 3 shows the two-factor model of the *MicroCog*, factor loadings, and factor correlation. Both the one and two-factor models fit the data quite well.

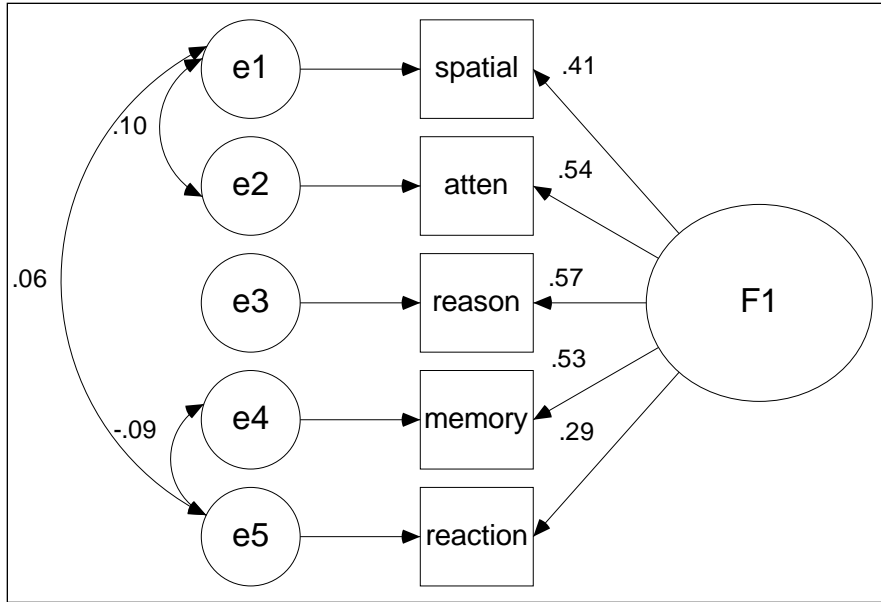


Figure 2 Factor Structure of the *MicroCog*

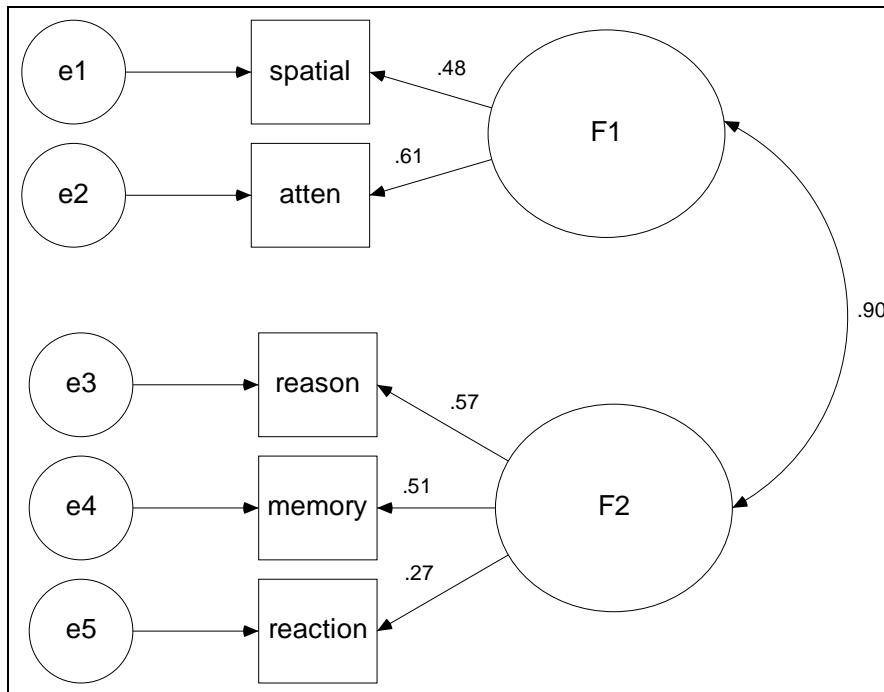


Figure 3 Factor Structure of *MicroCog* with Two Factors

The correlations of the factors in Figure 4 were all high and positive. This suggests a commonality of measured constructs.

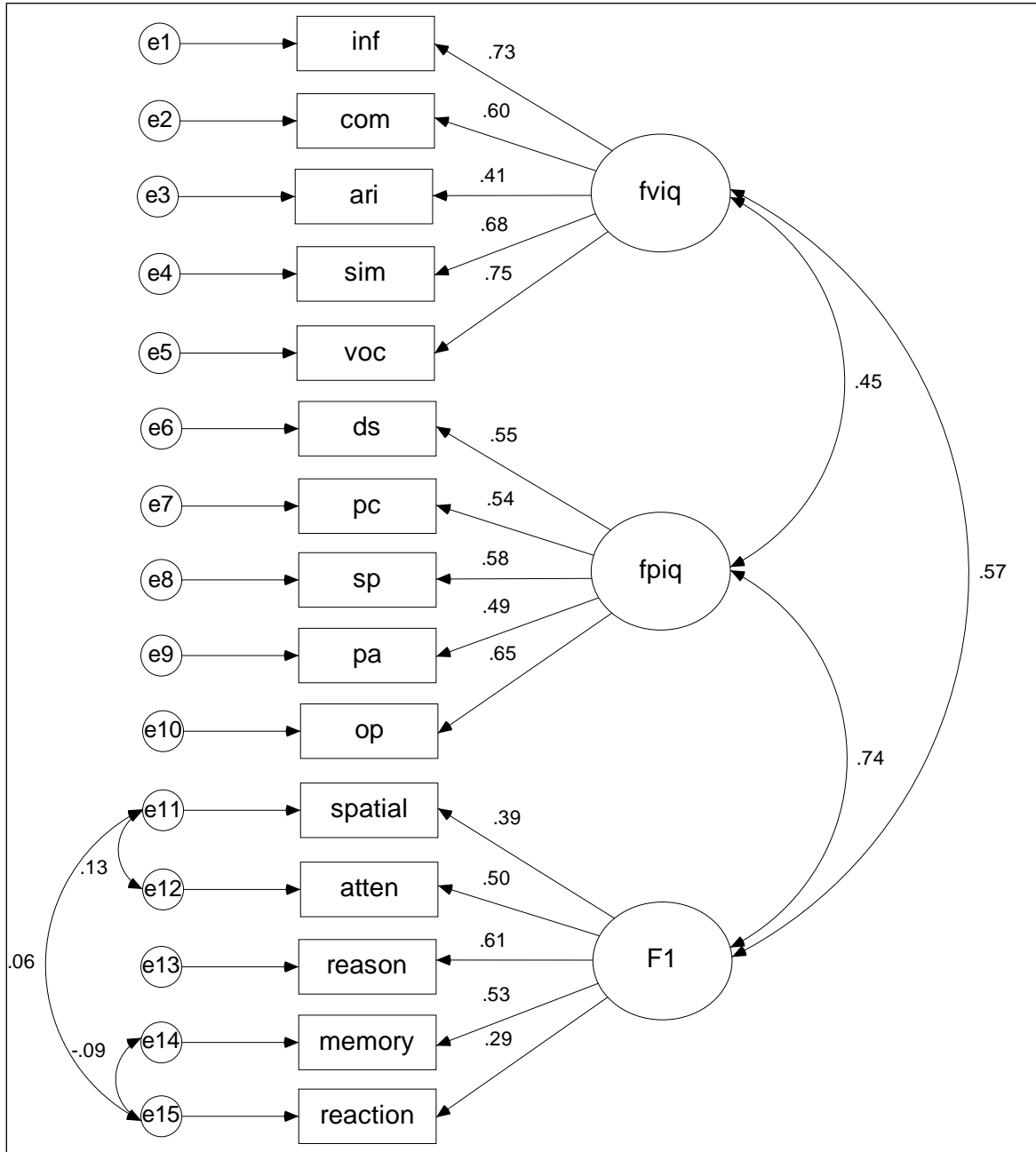


Figure 4 Joint Factor Structure of the MAB – II and MicroCog

DISCUSSION

The first purpose of this study was to compare the pilot trainee scores in this sample for both the *MAB – II* and the *MicroCog* to the nationally representative normative scores for each test.

The results of this study indicated that the FSIQ score on the *MAB - II* was substantially above the normative average, as were the Verbal and Performance IQ scores. The mean FSIQ was approximately 1.4 standard deviations above the norm, equivalent to the 92nd percentile. The mean PIQ and mean VIQ were both equivalent to the 90th percentile. The same was generally true of the *MicroCog*. The current Air Force student pilots are clearly above average in cognitive ability by all measures.

The second purpose was to compare mean scores of the current sample to those obtained in previous research. Over the last 20 years, the Air Force has been consistent in selecting pilots with high intellectual ability. Student pilots with high scores on ability tests are more likely to complete training (Olea & Ree, 1994; Ree & Carretta, 1996; Ree, Carretta, & Teachout, 1995) and learn to fly with fewer hours in training air craft (Duke & Ree, 1996) yielding substantial benefits to the Air Force. These results are consistent with previous studies assessing the intellectual performance and capabilities of USAF pilots and pilot applicants with respect to the normative population (Retzlaff, Callister, & King, 1999; Retzlaff & Gibertini, 1988).

The third purpose was to examine the similarity of the constructs measured by the two tests. If the two tests measured exactly or nearly the same constructs, they might be interchangeable in use.

Subtest Correlations Within the *MAB - II* and *MicroCog*

The following standards suggested by Cohen (1988) were adopted for interpretation of all correlations. Correlations were divided into three groups based on Cohen's *d*, a measure of effect size. Correlations categorized as: (a) small ($r = .10$ to $.23$) have effect sizes, *d*, of $.20$ to $.49$; (b) moderate ($r = .24$ to $.36$) have effect sizes of $.50$ to $.79$; and (c) large, ($r = .37$ or greater) have effect sizes of $.8$ or greater.

***MAB – II* subtest correlations.** Moderate to large correlations were found among the subtests that comprised the VIQ factor of the *MAB – II*. Similarly, moderate to large correlations were found among the subtests that make up the PIQ factor of the *MAB – II*. The largest correlations among the VIQ subtests were between Vocabulary and Information ($r = .58$) and between Vocabulary and Similarities ($r = .53$). The largest correlations among PIQ subtests were between Object Assembly and Picture Completion ($r = .41$) and between Object Assembly

and Spatial ($r = .40$). A review of correlations between VIQ and PIQ subtests revealed moderate correlations between Digit Symbol and Arithmetic ($r = .27$), Spatial Analyses and Arithmetic ($r = .28$), Picture Completion and Information ($r = .28$), Picture Completion and Similarities ($r = .24$). These findings suggest: (a) numerical reasoning may influence, to some degree, performance in areas regarding visual coding and visual-rotation aptitudes; and (b) general verbal reasoning and fund of knowledge may influence, to some degree, visual constructive reasoning and visual attention to detail. However, it is important to note most correlations between VIQ and PIQ subtests were small, providing supportive evidence of the construct validity for the *MAB – II* as established by the test developer.

***MicroCog* first-level index correlations.** Half of the correlations among the first-level index scores for the *MicroCog* were moderate, ranging from .24 to .31. This finding suggests that a single underlying cognitive structure (i.e., *g*) effects performance on 4 out of the 5 indices. However, correlations between Memory and Spatial Processing, as well as correlations between Reaction Time and all the other subtests are small, ranging from .08 to .21. These findings suggest psychomotor reaction time is relatively independent of the other cognitive aptitudes. In particular, a change in general psychomotor reaction time does not necessarily reflect changes in information processing speed for the other cognitive aptitudes. Furthermore, the Memory Index largely assesses verbal memory and does not necessarily assess for visual-spatial memory, which is captured by the Spatial Index.

Subtest Correlations Between the *MAB – II* and *MicroCog*

The Reasoning Index of the *MicroCog* had moderate to large correlations ($r = .24$ to $.37$) with four out of five of the PIQ subtests (Digit Symbol, Spatial Analyses, Picture Arrangement, and Object Assembly) and one (i.e., arithmetic) of the five VIQ subtests of the *MAB – II*. It appears that the aptitude of reasoning is significantly influenced by *Fluid* intelligence (which is the underlying construct of the visual-performance based aptitudes measured by subtests of the PIQ), as well as the Arithmetic subtest of the VIQ. This suggests that, for the most part, the Reasoning index of the *MicroCog* is influenced by a person's ability to draw inferences, and conceptual reasoning largely independent of acquired knowledge.

The Memory Index of the *MicroCog* had moderate correlations with all five of the VIQ subtests and only one PIQ subtest (Digit Symbol) from the *MAB – II*. Hence, the Memory Index would have a moderately positive relationship with items on the Digit Symbol subtest because such items measure an aspect of visual coding dependent upon memory. The finding that the Memory Index is highly correlated with all five subtests of the VIQ suggests the aptitude is influenced by a manifestation of *Crystallized* intelligence (i.e., acquiring new knowledge and understanding). Although crystallized intelligence is not generally equated with memory, it relies heavily upon mental processes of encoding, storage and retrieval of information commonly associated with memory (Horn & Cattell, 1966).

For the most part, the correlations between the two tests were small, indicating the subtests on the *MAB – II* and the first-level index scores on the *MicroCog* measured relatively independent aptitudes. However, all correlations of cognitive and neurocognitive aptitudes were positive, suggesting an overall underlying single cognitive structure (often described as *g*) common to all cognitive domains.

However, it is notable the lowest correlations between seven of the ten subtests of the *MAB – II* and all five indices of the *MicroCog* were found for Reaction Time on the *MicroCog*. This suggests psychomotor reaction time, measured by tapping speed with the index finger in response to a visual stimulus, is not as well related to the information processing speed and accuracy associated with other cognitive aptitudes.

Pattern Analyses Between *MicroCog* Indices and *MAB – II* Subtests

There were several correlations that appeared to be highly relevant to pattern analysis. Such correlations are meaningful for clinical interpretation of test scores and reflect the relationships between various aptitudes logically relevant to flying. First, the correlation between spatial subtests on the *MicroCog* and the *MAB – II* is relatively small ($r = .23$). Although the tests appear to have significant overlap in their descriptions, they measure diverging spatial aptitudes. The visual conceptual reasoning assessed by the *MAB – II* spatial subtest is independent of memory for spatial arrays and visuo-spatial analysis of clock faces assessed by the *MicroCog* Spatial Index. Second, a large correlation ($r = .37$) was found between Arithmetic assessed by the the *MAB – II* and Reasoning/Calculation assessed by the *MicroCog*. This may be due, in part, to the similarities between items assessing numerical reasoning on both tests. This finding suggests that evaluatees, who have deficits on the Reasoning/Calculation Index, are likely to experience difficulties with numerical reasoning and processing as measured by the Arithmetic subtest. However, the Reasoning/Calculation had moderate correlations with four out of the five *MAB – II* PIQ subtests, indicating such an index is more related to visual-perceptual reasoning than verbal reasoning. Third, a moderate correlation was found between the Digit Symbol and Arithmetic subtests on the *MAB – II* and the Attention/Mental Control Index on the *MicroCog* ($r = .31$ and $.26$, respectively). It appears that declines in visual working memory and learning and numerical processing may also be manifested in difficulties sustaining attention and mental control (i.e., distraction). Fourth, a moderate correlation was found between the Similarities subtest on the *MAB – II* and the Memory Index of the *MicroCog* ($r = .29$). This would indicate there is a positive relationship between verbal reasoning and knowledge of verbal concepts and general verbal memory. The items on the Memory Index primarily assess verbal aspects due to item composition that is largely based on memory for addresses.

Confirmatory Factor Analysis

The model for the *MAB – II* fit the data quite well even though the data were range restricted. A difference was found between the one and two-factor *MicroCog* measurement models. The one-factor model fit the data significantly better than the two-factor model with all goodness-of-fit indexes at or above the minimum levels indicating a good fit (GFI \geq .9, CFI \geq .9, RMSEA \leq .08, CN \geq 200).

Examination of the three factors derived from the joint *MAB – II* and *MicroCog* confirmatory factor analysis shows a stronger correlation between the single *MicroCog* factor and the Performance IQ factor ($r = .74$) than the Verbal IQ factor ($r = .57$). This suggested that although the *MAB – II* and the *MicroCog* measure much in common, there is a stronger correlation among visual-performance based aptitudes.

The average correlation among the subtests of the *MAB – II* was higher than the average correlations of the *MicroCog* suggesting that more of the score variance was due to the measurement of common variance, g . The *MAB – II* has been shown to have two correlated factors. The common variance in the first factor (VIQ) is 52% and 45% in the second factor (PIQ). However, the common variance in the *MicroCog* was 38%. All three of these common variances can be expected to be greater in a sample that was not subject to prior selection causing range restriction. In most test batteries, the common variance (g), usually accounts for about half or more of the variability in the scores (Carretta & Ree, 1996; Ree & Carretta, 1994).

Aeromedical Case Study Application

Interpretation of psychological test data involves the use of normative data and pattern analysis. Through the use of normative data it is possible to view the strengths and direction of an individual's cognitive aptitudes as compared with peers. Pattern analysis, on the other hand, involves examining the person's scores to develop an overall profile that provides an indication of how adaptive a person's cognitive aptitudes are in various settings and situations. For example, if a pilot applicant's scores on cognitive testing are low in areas such as visual memory and learning, spatial processing, and visual construction (when compared with other pilots), then there would be reason to suspect the person's cognitive aptitudes may be incompatible with the inherent visual-performance-based demands of military flying. On the other hand, for a pilot who has been disqualified from flying due to a history of an organic condition (e.g., bacterial meningitis) that temporarily effected his or her cognitive functioning (e.g., attention, memory), but whose scores now reflect the absence of cognitive difficulties, and is within normal limits when compared with other pilots, it may be reasonable to conclude the pilot could return to flying duties.

It is important to utilize the data presented in Tables 1- 5 in this study when clinically evaluating and interpreting the scores of a pilot or pilot candidate. The percentile tables allow a psychologist to determine how a specific score compares with the distribution of scores for rated USAF pilots as a group. In general, scores that fall above the 90th or below the 10th percentile can be reasonably viewed as outliers and significantly different from most rated AF pilots.

The utility of these data is illustrated by the case of an F-16 fighter pilot with a history of a mild head injury during a soccer game. He had collided with another player while attempting to score a goal. During the collision he bounced off the other player and inadvertently struck the right parietal section of his head against the goal post. When he struck his head, witnesses reported he lost consciousness for about 15 - 30 seconds. He reported having only partial memory of the incident and cannot remember much of the hour following the incident. He reported feeling “dazed” and “nauseated” for a few minutes after the injury with a moderate headache that persisted for a few hours. He was taken to the hospital and imaging did not reveal any sort of epidural bleeding or parenchymal injury. He denied having any additional sequela and stated the only symptom that persisted was a sensation of tenderness on the right side of his head that resolved within a week.

The pilot’s history of a mild closed head injury disqualifies him from flying according to USAF aeromedical policy (AFI 48-123, 2009). In this case, a neuropsychological screening was requested following injury to assess the presence of any changes in the pilot’s cognitive disposition. The pilot was referred to the installation’s active-duty clinical psychologist for an evaluation and screening. The psychologist who evaluated the pilot included computer based intelligence testing (i.e., *MAB – II*) and neuropsychological aptitude screening (i.e., *MicroCog*) in the evaluation. It was particularly important for the psychologist to determine if: (a) the history of a head injury continued to affect the pilot’s cognitive functioning a month after the incident; and (b) the pilot’s cognitive disposition was compatible with managing the rigorous nature of a pilot’s duties. See Table 8 for a case example of post-injury scores as compared with baseline testing.

The pilot’s post-injury responses to items on the *MAB – II* (when compared with males in the civilian, non-pilot normative sample) were well within normal limits. For example, according to the data from Table 8, the pilot’s Digit Symbol, Spatial, Object Assembly *MAB-II* scores and the Attention/Mental Control, Reasoning, and Spatial Index scores on the *MicroCog* were within the average to high-average range when compared to a population with similar age and education. To many psychologists, it would be reasonable to conclude there is no change in his cognitive disposition and that his scores are within normal limits. However, it would have been incorrect to conclude the pilot’s scores were within normal limits for *USAF pilots*. When utilizing the descriptive data from Tables 4 and 5, such scores are at (or well below) the 10th percentile when compared with USAF pilots.

Table 8. Case Example: Baseline and 1-Month Post-injury Performance on the *MAB – II* and *MicroCog*.

<u>Test</u>	<u>General Population Comparison</u>		<u>USAF Pilot Comparison</u>	
	Baseline	Post Injury	Baseline	Post Injury
<i>MAB-II</i> T-scores (age corrected)				
Information	68 (96%)	63 (91%)	68 (50%)	63 (25%)
Comprehension	61 (86%)	61 (86%)	61 (50%)	61 (50%)
Arithmetic	58 (79%)	52 (58%)	58 (25%)	52 (10%)*
Similarities	57 (75%)	61 (50%)	57 (25%)	61 (50%)
Vocabulary	60 (84%)	60 (84%)	60 (50%)	60 (50%)
Digit Symbol	69 (97%)	55 (70%)	69 (50%)	55 (5%)*
Picture Completion	65 (94%)	61 (87%)	65 (75%)	61 (50%)
Spatial	61 (87%)	52 (55%)	61 (50%)	52 (10%)*
Picture Arrangement	58 (79%)	51 (53%)	58 (75%)	51 (50%)
Object Assembly	59 (83%)	53 (61%)	59 (25%)	53 (5%)*
IQ Standard Scores				
Verbal IQ	117 (87%)	114(83%)	117 (39%)	114(24%)
Performance IQ	119 (90%)	107(68%)	119 (51%)	107(10%)*
Full scale IQ	118 (88%)	111(77%)	118 (37%)	111(10%)*
<i>MicroCog</i> First-Level Indexes (age and education corrected)				
Attention/Mental Control	104 (53%)	87 (19%)	104 (50%)	87 (10%)*
Reasoning/Calculation	98 (45%)	80 (10%)	98 (50%)	80 (10%)*
Memory	101 (53%)	100 (50%)	101 (25%)	100 (23%)
Spatial Processing	114 (83%)	93 (32%)	114 (75%)	93 (10%)*
Reaction Time	106 (66%)	100 (50%)	106 (75%)	100 (50%)

Note: * indicates post-injury scores at the 10th percentile or lower compared to pilot norms

In addition to having an adequate understanding of how the pilot's current scores are related to USAF pilot norms, it is equally, if not more important to compare such scores with baseline testing obtained during medical flight screening three years earlier. The results of

Table 8 indicate the pilot's baseline testing scores were in the high-average to superior range and more consistent with USAF pilot norms. However, the post-injury scores reflect a noticeable change in the pilot's performance, mainly in the areas of visual-spatial aptitudes, indicating his cognitive disposition had not fully returned to baseline. It is important to note that testing occurred under the same standardized physical and environmental conditions and motivational level to perform well.

The post-injury cognitive profile revealed decline in the areas of visual spatial perception, figural coding, visual constructive reasoning, as well as numerical reasoning on the *MAB – II*. This corresponded to declines in functioning associated with memory for novel spatial arrays, attention/vigilance to visual stimuli, as well as figural and numerical reasoning on the *MicroCog*. Such visual-spatial and perceptual aptitudes are considered “no hold” tests that measure aspects of fluid intelligence likely to reflect cognitive decline due to medical and psychological illness/injury. The results indicated his head injury was more severe than a “mild” classification, and a more comprehensive neuropsychological evaluation is needed due to the objective data reflecting a decline in such visual-spatial functioning.

Although the changes in his cognitive aptitudes may lead to several different hypotheses and speculation over the region of brain dysfunction, the pattern of his scores is suggestive of parietal lobe dysfunction and “axonal shearing” not readily identified via imagery. The polymodal region of the posterior parietal cortex is important in various aspects of “spatial cognition” related to the mental rotation and manipulation of visual images, nonverbal memory, as well as numerical reasoning. Furthermore, the changes in his cognitive aptitude profile are consistent with posterior parietal injury.

After reviewing additional information, the clinical neuropsychologist concluded the pilot's cognitive functioning was lower than expected. His visual-performance based aptitudes had not fully returned to baseline and were significant outliers when compared with other USAF pilots. The decline in his cognitive performance raised concerns about the compatibility of his current cognitive disposition with the challenging and dangerous conditions associated with military flying of a single seat jet fighter aircraft. As a result, the cognitive functioning of the pilot did not appear to meet the aeromedical criterion for flying, and it was believed that the pilot could benefit from additional observation, more in-depth neuropsychological testing, and a greater period of healing time prior to resumption of flying duties. The normative data included in Tables 4 and 5 helped the aeromedical psychologist to accurately interpret the pilot's test scores in comparison to USAF pilots as a group, and to render a difficult and very important decision about the pilot's readiness to return to fly, where the safety risks and potential losses from an aviation mishap are high. If USAF pilot norms and baseline testing had not been used, the pilot may have been cleared to return to fly in a cognitively impaired state.

LIMITATIONS TO THE 'BEST FIT MODEL'

Although the correlational matrix and normative scores are essential to aeromedical interpretation of pilot and pilot applicant test scores, establishing a 'best fit' factorial model is a more complex process. The factorial model in this study is based upon several steps that involve the following.

First, the *MAB-II* was divided into two factors based upon the manufacturers guidelines. These two factors consisted of a Verbal (inf, com, ari, sim, voc) and Performance (ds, pc, sp, pa, op) components as established by the test publisher. This two factor model is based upon administration and factorial analysis of the test battery with a 'normal' population.

Second, the *MicroCog* is considered a 'one factor' model with five 'Indices' that include items assessing: Spatial Analyses, Attention/Mental Control, Reasoning/Calculation, Memory and Reaction Time. Two separate factor models were considered. A one factor model and two factor model, respectively. The two factor model consisted of dividing the Indices into Factor 1 (Spatial, Attention) and Factor 2 (Reason, Memory, Reaction Time).

Third, a Structural Equation Confirmatory Factor Analysis model was fit to the data using the one factor *MicroCog* and the two factor *MAB-II* model. Based upon established criteria from the Structural Equation Literature the following indices were used to assess a 'Best Fit' model: (a) Comparative Fit Index (CFI) >.90; (b) Goodness of Fit Index (GFI) >.90; (c) Root Mean Square Error Approximation (RMSEA) <.05; (d) Hoelter N known as Critical N >200; and (e) Modification Indices all <100.

Based upon the above criteria the best fit model was a one factor *MicroCog* and two factor *MAB-II*. Many other indices (e.g., correlation scores between subtests and Indices) were used to further assess the appropriateness of this fit. While this methodology is sound and provides one possible answer to the question of the relationship between the *MAB-II* and *MicroCog*, there may yet be the possibility that an additional approach might provide a better model fit when giving consideration to the 'uniqueness' of the USAF pilot population. Because this population demonstrates certain 'unique' skill sets, it is possible the USAF pilot population 'does not' respond to the *MAB-II* and *MicroCog* in the same way as the normal population.

It should also be mentioned that, while the 'best fit model' establishes significant relationships among the various variables of both the *MAB-II* and *MicroCog*, the model does not suggest predictive success. The model must be subjected to predictive modeling to determine if the variables can accurately predict pilot training outcomes.

Therefore, the 'best fit model' in this study is only designed to establish a proof of concept and not represent an exhaustive approach to modeling a predictive psychological profile of cognitive aptitudes most predictive of success or failure in pilot training. Further modeling must be conducted to refine the predictive model based upon the relationships among the variables of the *MAB-II* and *MicroCog*.

CONCLUSION

Population specific normative data are needed to effectively evaluate rated USAF pilots when rendering aeromedical decisions about their psychological disposition. This is especially the case when assessing how the psychological disposition of a pilot applicant compares with current USAF pilots. There are significant differences between the general population and rated USAF pilot normative scores. As a result, psychologists should be sensitive to such differences and utilize population specific normative data that more accurately represent the group with which a person is being compared. Otherwise, as in the case illustrated above, a psychologist may mistakenly conclude that the given scores were within normal limits when, in fact, the scores were not when compared with population specific norms (i.e., norms for rated pilots). Appropriate normative data are central to the clinical interpretation of psychological test scores. These are regularly a part of the evaluation process for selecting applicants for pilot training, and for considering a pilot's readiness for returning to flying duties after being disqualified for psychological reasons. Furthermore, it is clear that performance on intelligence testing with the *MAB-II* and neuropsychological screening with the *MicroCog*, are related. There is a stronger correlation between the single factor of the *MicroCog* and Performance IQ of the *MAB – II* than between the single factor of the *MicroCog* and the Verbal IQ of the *MAB – II*. Further, it is clear that the *MAB – II* and *MicroCog* are not interchangeable, given the less than $r = 1.0$ correlations among the factors.

REFERENCES

- Air Force Instruction (AFI) 48-123 (2009). Aerospace Medicine: Medical Examinations and Standards. Retrieved October 5, 2009 from <http://www.e-publishing.af.mil/forms-pubs/>
- Arbuckle, J.L. (2007). *Amos 16.0 User's Guide*. Chicago, IL: SPSS Incorporated.
- Blunch, N. (2008). *Introduction to Structural Equation Modeling Using SPSS and Amos*. Thousand Oaks, CA: Sage.
- Callister, J., King, R., & Retzlaff, P. (1996). Cognitive assessment of USAF pilot training candidates. *Aviation, Space, and Environmental Medicine*, 67, 1124 – 1129
- Carretta, T.R., & Ree, M.J. (1996). Factor structure of the Air Force Officer Qualifying Test: Analysis and comparison. *Military Psychology*, 8, 29-42.
- Carretta, T.R., Retzlaff, P.D., & King, R.E. (1997). *A tale of two test batteries: A comparison of the Air Force Officer Qualifying Test and the Multidimensional Aptitude Battery (AL/HR-TP-1997-0052)*. Mesa, AZ: United States Air Force Armstrong Laboratory, Human Resources Directorate.
- Carretta, T.R., Retzlaff, P., Callister, J., & King, R. (1998). A comparison of two U.S. Air Force pilot tests. *Aviation, Space, and Environmental Medicine*, 69, 931 – 935.
- Christal, R.E. (1990). *Comparative validities of ASVAB and LAMP tests for logic-gate learning*. Unpublished manuscript. Brooks Air Force Base, Manpower and Personnel Division, Air Force Human Resources Laboratory, Texas.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). New Jersey: Lawrence Erlbaum.
- Duke, A.P., & Ree, M.J. (1996). Better candidates fly fewer training hours: Another time testing pays off. *International Journal of Selection and Assessment*, 4, 115-121.
- Horn, J. L., & Cattell, R. B. (1966). Refinement and test of the theory of fluid and crystallized general intelligences. *Journal of Educational Psychology*, 57, 253-270.
- Jackson, D.N. (1984). *Multidimensional Aptitude Battery - Manual*. London: Research Psychologists Press.
- Jackson, D.N. (1998). *MAB – II Multidimensional Aptitude Battery*. Lutz, FL: Psychological Assessment Resources, Inc.

Jackson, D.N. (2003). *Multidimensional Aptitude Battery-II: Manual*. Port Huron, MI: Sigma Assessment Systems, Inc.

King, R. & Flynn, C.F. (1995). Defining and measuring the "Right Stuff": Neuropsychiatrically Enhanced Flight Screening (N-EFS). *Aviation, Space, and Environmental Medicine*, 66, 951 – 956.

Kranzler, J.H. (1991). The construct validity of the Multidimensional Aptitude Battery: A word of caution. *Journal of Clinical Psychology*, 47, 691 – 697.

Kyllonen, P.C. (1985). *Dimensions of information processing speed* (Technical Paper AFHRL-TP-84-56). Brooks Air Force Base, Manpower and Personnel Division, Air Force Human Resources Laboratory, Texas.

Kyllonen, P.C. (1991). CAM: A theoretical framework for cognitive abilities measurement. D. Detterman (Ed.), *Current Topics in Human Intelligence: Theories of Intelligence*. Norwood, NJ: Ablex.

Lance, C.E., & Vandenberg, R.J. (2009). *Statistical and methodological myths and urban legends: Doctrine, verity and fable in the organizational and social sciences*. New York: Routledge.

Lee, M.S., Wallbrown, F.H., & Blaha, J. (1990). Note on the construct validity of the Multidimensional Aptitude Battery. *Psychological Reports*, 67, 1219 – 1222.

Neudecker, H., & Magnus, J.R. (1988). *Matrix differential calculus with applications in statistics and econometrics*. New York, NY: John Wiley.

Olea, M., & Ree, M.J. (1994). Predicting pilot and navigator criteria: Not much more than *g*. *Journal of Applied Psychology*, 79, 845-851.

Powell, D.H., Kaplan, E.F., Whitla, D., Weintraub, S., Caitlin, R., & Funkenstein, H.H. (1993). *MicroCog: Assessment of Cognitive Functioning Manual*. San Antonio, TX: Psychological Corporation.

Ree, M.J., & Carretta, T.R. (1994). Factor analysis of the ASVAB: Confirming a Vernon-like structure. *Educational and Psychological Measurement*, 54, 457-461.

Ree, M.J., & Carretta, T.R. (1996). Central role of *g* in the prediction of pilot selection. *The International Journal of Aviation Psychology*, 6, 111-123.

Ree, M.J., Carretta, T.R., & Teachout, M.S. (1995). Roll of ability, and prior knowledge in complex training and performance. *Journal of Applied Psychology*, 80, 721-730.

Ree, M.J., & Earles, J.A. (1991). The stability of convergent estimates of g. *Intelligence, 15*, 271-278.

Retzlaff, P., Callister, J., & King, R. (1999). Clinical procedures for the neuropsychological evaluation of U. S. Air Force pilots. *Military Medicine, 164*, 1-6.

Retzlaff, P., & Gibertini, M. (1988). Objective psychological testing of U. S. Air Force officers in pilot training. *Aviation, Space, and Environmental Medicine, 59*, 661-663.

Retzlaff, P., King, R., & Callister, J. (1995). *U.S. Air Force pilot training completion and retention: A ten year follow-up on psychological testing*. (AL/AO-TR-1995-0124). Brooks AFB: Armstrong Laboratory.

Wallbrown, F.J., Carmin, C.N., & Barnett, R.W. (1988). Investigating the construct validity of the Multidimensional Aptitude Battery. *Psychological Reports, 62*, 871 – 878.

Wallbrown, F.J., Carmin, C.N., & Barnett, R.W. (1989). A further note on the construct validity of the Multidimensional Aptitude Battery. *Journal of Clinical Psychology, 45*, 429 – 433.

Wechsler, D. (1997). *Wechsler Adult Intelligence Scale – 3rd Edition (WAIS – 3 ®)*. San Antonio, TX: Harcourt Assessment.