



**AFRL-RH-WP-TR-2014-0014**

**JOINT ANALYSIS OF TWO ABILITY TESTS:  
TWO THEORIES, ONE OUTCOME**

Mark S. Teachout  
University of the Incarnate Word  
San Antonio, TX

Malcolm James Ree    Erica L. Barto  
Operational Technologies, Corporation  
San Antonio, TX

Thomas R. Carretta  
Air Force Research Laboratory  
Wright-Patterson AFB, OH

Raymond E. King  
Headquarters, Air Force Safety Center  
Kirtland AFB, NM

**January 2014  
Interim Report**

**Distribution A: Approved for public release; distribution unlimited.**

**AIR FORCE RESEARCH LABORATORY  
711 HUMAN PERFORMANCE WING,  
HUMAN EFFECTIVENESS DIRECTORATE,  
WRIGHT-PATTERSON AIR FORCE BASE, OH 45433  
AIR FORCE MATERIEL COMMAND  
UNITED STATES AIR FORCE**

## NOTICE AND SIGNATURE PAGE

Using Government drawings, specifications, or other data included in this document for any purpose other than Government procurement does not in any way obligate the U.S. Government. The fact that the Government formulated or supplied the drawings, specifications, or other data does not license the holder or any other person or corporation; or convey any rights or permission to manufacture, use, or sell any patented invention that may relate to them.

Qualified requestors may obtain copies of this report from the Defense Technical Information Center (DTIC) (<http://www.dtic.mil>).

AFRL-RH-WP-TR-2014-0014 HAS BEEN REVIEWED AND IS APPROVED FOR PUBLICATION IN ACCORDANCE WITH ASSIGNED DISTRIBUTION STATEMENT.

//signed//

Thomas R. Carretta  
Engineering Research Psychologist  
Supervisory Control and Cognition Branch  
Decision Making Division

//signed//

Gregory J. Barbato  
Chief, Supervisory Control and Cognition Branch  
Decision Making Division

//signed//

William E. Russell  
Chief, Decision Making Division  
Human Effectiveness Directorate  
711 Human Performance Wing

This report is published in the interest of scientific and technical information exchange, and its publication does not constitute the Government's approval or disapproval of its ideas or findings.

# REPORT DOCUMENTATION PAGE

*Form Approved*  
*OMB No. 0704-0188*

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

<b>1. REPORT DATE (DD-MM-YYYY)</b> 30-01-2014		<b>2. REPORT TYPE</b> Interim		<b>3. DATES COVERED (From - To)</b> 03 Jan 2013 – 30 Jun 2013	
<b>4. TITLE AND SUBTITLE</b>  Joint Analysis of Two Ability Tests: Two Theories, One Outcome				<b>5a. CONTRACT NUMBER</b> NA	
				<b>5b. GRANT NUMBER</b>	
				<b>5c. PROGRAM ELEMENT NUMBER</b> 62202F	
<b>6. AUTHOR(S)</b>  Mark S. Teachout <sup>1</sup> , Malcolm James Ree <sup>2</sup> , Erica L. Barto <sup>2</sup> , Thomas R. Carretta <sup>3</sup> , and Raymond E. King <sup>4</sup>				<b>5d. PROJECT NUMBER</b> 5329	
				<b>5e. TASK NUMBER</b> 09	
				<b>5f. WORK UNIT NUMBER</b> H03K/53290902	
<b>7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)</b> See next page.				<b>8. PERFORMING ORGANIZATION REPORT NUMBER</b> N/A	
<b>9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)</b> Air Force Materiel Command Air Force Research Laboratory 711 Human Performance Wing Human Effectiveness Directorate Warfighter Interface Division Supervisory Control and Cognition Branch Wright-Patterson AFB OH 45433				<b>10. SPONSOR/MONITOR'S ACRONYM(S)</b> 711 HPW/RHCI	
				<b>11. SPONSOR/MONITOR'S REPORT NUMBER(S)</b> AFRL-WP-RH-TR-2014-0014	
<b>12. DISTRIBUTION / AVAILABILITY STATEMENT</b> Distribution A: Approved for public release; distribution is unlimited.					
<b>13. SUPPLEMENTARY NOTES</b> 88 ABW Cleared 03/20/2014; 88ABW-2014-1128.					
<b>14. ABSTRACT</b> Distinctions between tests of general cognitive ability ( <i>g</i> ) versus specific abilities ( <i>s</i> ) have been investigated for over a century. The similarity of two tests designed using two theories was examined. The Multidimensional Aptitude Battery used the individual differences psychometric approach while the MicroCog used a brain-behavior relationships approach. Both tests were administered to 10,612 participants. Correlations suggested the tests shared a common source of variance for the constructs measured. Confirmatory factor analyses established this and a hierarchical structure with <i>g</i> at the apex. The test designed to measure specific abilities (MicroCog) measured a single factor, <i>g</i> . Although different theories underlie their respective construction methods, results indicated the two tests measured much in common. That is, two theories, one outcome.					
<b>15. SUBJECT TERMS</b> General cognitive ability, specific abilities, confirmatory factor analysis, test development theory, indifference of the indicator					
<b>16. SECURITY CLASSIFICATION OF:</b>			<b>17. LIMITATION OF ABSTRACT</b>  SAR	<b>18. NUMBER OF PAGES</b>  27	<b>19a. NAME OF RESPONSIBLE PERSON</b> Antonio Ayala
<b>a. REPORT</b> Unclassified	<b>b. ABSTRACT</b> Unclassified	<b>c. THIS PAGE</b> Unclassified			<b>19b. TELEPHONE NUMBER (include area code)</b>

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

7. PERFORMING ORGANIZATIONS NAMES AND ADDRESSES (cont.):

Mark. S. Teachout<sup>1</sup>  
University of the Incarnate Word  
4301 Broadway, CPO491  
San Antonio, TX 78209

Malcolm James Ree<sup>2</sup> Erica L. Barto<sup>2</sup>  
Operational Technologies, Corporation  
4100 N. W. Loop 410  
San Antonio, TX 78229

Thomas R. Carretta<sup>3</sup>  
Air Force Research Laboratory  
711 HPW/RHCI  
2210 8<sup>th</sup> Street  
Area B, Bldg 146, Room 122  
Wright-Patterson AFB, OH 45433

Raymond E. King<sup>4</sup>  
Headquarters, Air Force Safety Center  
AFSEC/SEH  
Bldg. 24499  
Kirtland AFB, NM 85117

# TABLE OF CONTENTS

SECTION	Page
<b>Table of Contents</b> .....	<b>iii</b>
<b>Lof Tables</b> .....	<b>iv</b>
<b>List of figures</b> .....	<b>iv</b>
<b>Preface</b> .....	<b>v</b>
<b>1.0 SUMMARY</b> .....	<b>1</b>
<b>2.0 INTRODUCTION</b> .....	<b>1</b>
2.1 Test Construction Approaches .....	1
2.1.1 Individual Differences Psychometric Approach .....	1
2.1.2 Specific Abilities approach .....	2
2.2 Joint Relationships of Two Measures of Ability .....	3
2.3 Purpose .....	3
<b>3.0 METHODS</b> .....	<b>4</b>
3.1 Participants .....	4
3.2 Measures .....	4
3.2.1 MAB-II .....	4
3.2.2 MicroCog .....	5
3.3 Procedures .....	6
<b>4.0 RESULTS</b> .....	<b>7</b>
4.1 Corrected Correlations .....	9
4.1.1 MAB-II .....	9
4.1.2 MicroCog .....	9
4.1.3 MAB-II and MicroCog .....	9
4.2 Confirmatory Factor Analysis .....	9
<b>5.0 DISCUSSION</b> .....	<b>11</b>
<b>6.0 REFERENCES</b> .....	<b>16</b>
<b>LIST OF ACRONYMS, ABBREVIATIONS, AND SYMBOLS</b> .....	<b>19</b>

## LIST OF TABLES

<b>Table</b>		<b>Page</b>
1	MAB-II Factors, Subtests, and Descriptions .....	5
2	MicroCog Index Descriptions.....	6
3	Means, Standard Deviations, and Correlations of MAB-II subtests and MicroCog First-Level Indices.....	8
4	Goodness-of-Fit Statistics for the Two Measurement Models.....	9

## LIST OF FIGURES

<b>Figure</b>		<b>Page</b>
1	MAB-II Confirmatory Factor Model .....	10
2	MicroCog Confirmatory Factor Model.....	11
3	Joint Factor Structure of the MAB-II and MicroCog .....	12
4	Apex of the Hierarchical Model for the MAB-II and MicroCog .....	13

## **PREFACE**

This report describes activities performed during the examination of historical data regarding measures of US Air Force pilot candidate aptitude and training outcomes (711 HPW/RHCI), Work Unit 53290902.

The opinions expressed are those of the authors and not necessarily those of the United States Government, Department of Defense, or the United States Air Force. Address all correspondence to Dr. Mark Teachout at [teachout@uiwtx.edu](mailto:teachout@uiwtx.edu).

**THIS PAGE INTENTIONALLY LEFT BLANK**

## 1.0 SUMMARY

Distinctions between tests of general cognitive ability (*g*) versus specific abilities (*s*) have been investigated for over a century. The similarity of two tests designed using two theories was examined. The Multidimensional Aptitude Battery used the individual differences psychometric approach while the MicroCog used a brain-behavior relationships approach. Both tests were administered to 10,612 participants. Correlations suggested the tests shared a common source of variance for the constructs measured. Confirmatory factor analyses established this and a hierarchical structure with *g* at the apex. The test designed to measure specific abilities (MicroCog) measured a single factor, *g*. Although different theories underlie their respective construction methods, results indicated the two tests measured much in common; that is, two theories, one outcome.

## 2.0 INTRODUCTION

For over a century, a major area of research has involved the distinction between tests that measure general cognitive ability (*g*) and tests designed to measure specific abilities (*s*). The measurement of general and specific abilities reflects two different test construction approaches driven by different theories. These approaches should produce results consistent with each underlying theory, yet different from one another. The constructs measured should reflect these differences. Research on traditional cognitive ability tests developed using an individual differences psychometric approach, has shown that they have a hierarchical structure with *g* at the apex and are good predictors of occupational performance. There has been extensive research on tests intended to measure specific abilities (*s*) constructed using different underlying theoretical approaches. Results suggest that when examined for *g*, these specific abilities approaches also largely measure *g* (Murphy, 2009). Further, *s* adds little to the prediction of occupational performance beyond *g* (e.g., Ree & Earles, 1991; Ree, Earles, & Teachout, 1994). The current study examined two tests designed using different underlying theoretical approaches to determine the extent to which each test measured what it purports to measure, general vs. specific abilities. These two approaches were the traditional individual differences psychometric approach, exemplified by the Multidimensional Aptitude Battery (MAB-II; Jackson, 1998) and an approach based on physiological theories of brain-behavior relationships (Kaplan, 1988), exemplified by the MicroCog (Powell, Kaplan, Whitla, Weintraub, & Funkenstein, 1993).

### 2.1 Test Construction Approaches

**2.1.1 Individual Differences Psychometric Approach.** The traditional individual differences psychometric approach is based on decades of well-established research in ability testing. Tests are developed across a broad range of content (e.g., verbal, math, perceptual) and are intended to measure the abilities represented by the content domains. The information is used to make

distinctions among individuals. While this approach was not designed with the intent of measuring general cognitive ability, research has shown that tests designed with this approach measure these content domains and contain a substantial measure of *g* (Jensen, 1998).

Galton (1869) proposed a general ability factor, later operationalized by Spearman (1904, 1927) in his two-factor theory of ability. The first factor was general cognitive ability (*g*); the second was specific (*s*) to each test, measuring a specific content. Since Spearman's (1904, 1927) early work, almost a century of research has confirmed the nature and validity of cognitive ability tests. For example, Vernon (1950, 1969) demonstrated the hierarchical structure of multiple aptitude batteries, with *g* at the apex. Numerous studies have confirmed the structure of cognitive ability, demonstrating the pervasiveness of psychometric *g* (e.g., Jensen, 1998; Ree & Carretta, 1994). The validity of *g* has been well documented as the best predictor of academic performance in 4,096 studies conducted with the Differential Aptitude Test (McNemar, 1964) and for numerous occupational criteria where meta-analyses found that the primary source of validity is *g* (Hunter & Hunter, 1984). Specific cognitive abilities have added little to the validity of ability tests beyond *g* (Jensen, 1980, 1998; McNemar, 1964; Ree & Earles, 1991, 1992; Ree et al., 1994; Ree, Teachout, & Carretta, 2013).

**2.1.2 Specific Abilities Approach.** In contrast to the individual differences psychometric approach, the specific abilities approach is designed to *intentionally* produce distinct measures of specific abilities. Tests are developed within a narrow range of specific domains congruent with specific content, processes, or cognitive components. Specific cognitive contents include verbal, math, spatial, perceptual speed, reasoning, and mechanical. Specific cognitive processes include stages of information processing, such as acquisition, processing, and retrieval. Cognitive components include declarative and procedural knowledge, working memory, and information processing speed. Subtests are developed and administered to measure the abilities represented by each specific domain, and the information is used to make distinctions among individuals. Unlike the individual differences approach, specific ability theories attempt to measure and demonstrate the unique existence of specific abilities (e.g., Kyllonen, 1994).

Early challenges to *g* have included Hull (1928) and Thurstone (1938) who proposed that specific abilities could compensate for *g*. Between WWI and WWII, many theorists shifted to a multiple abilities perspective, giving rise to multiple aptitude batteries (e.g., Air Force Officer Qualifying Test, Armed Services Vocational Aptitude Battery, Differential Aptitude Tests, General Aptitude Test Battery). While these batteries were constructed with the intent to measure specific abilities, subsequent research indicated they contained substantial amounts of *g* (Jensen, 1998).

Recent approaches have been used to develop tests of specific abilities. Kyllonen (1993, 1994) focused on the measurement of cognitive components (i.e., declarative knowledge, processing

speed, procedural knowledge, and working memory). A neuropsychological approach based on theories of brain function and underlying brain-behavior relationships emphasizes the distinction between different cognitive domains and their underlying cortical networks (Powell et al., 1993).

## 2.2 Joint Relationships of Two Measures of Ability

Several studies have examined relationships between two ability tests using the individual differences psychometric approach. Sperl, Ree, and Steuck (1992) examined the relations between the verbal and math tests of two military multiple aptitude batteries, the Air Force Officer Qualifying Test (AFOQT) and the Armed Services Vocational Aptitude Battery (ASVAB), finding a canonical correlation of .93, indicating a high level of common variance. Carretta, Retzlaff, Callister, and King (1998) studied the relationship between the hierarchical factors of two multiple aptitude batteries, the MAB-II and the AFOQT, finding that both tests had a higher order factor (*g*), with a correlation of .98 between the higher order factors. In a joint confirmatory factor analysis (CFA) of the Differential Abilities Scales and the Woodcock-Johnson Tests of Cognitive Abilities, Sanders, McIntosh, Dunham, Rothlisberg, and Finch (2007) demonstrated a hierarchical structure with *g* at the apex. Results from these studies indicated that psychometric *g* underlies these ability tests.

Few studies have investigated the joint factor structure of measures of ability comparing the individual differences psychometric approach to the specific abilities approach. Kyllonen (1993) examined the relationships between the ASVAB, a multiple abilities battery, and a battery of specific cognitive components. He found that working memory was related to *g*, and asserted that working memory is *the* general factor in cognition. Extending Kyllonen (1993), Stauffer, Ree, and Carretta (1996) examined the common sources of variance across the two tests and found that both measured *g*, with the higher order factors correlated at .99. Similarly, Keith, Kranzler, and Flanagan (2001) compared two tests using differing theoretical approaches. The Woodcock-Johnson III Tests of Cognitive Abilities is an intelligence test based on the individual differences psychometric theory (Woodcock, McGrew, & Mather, 2001). The specific test battery was the Cognitive Assessment System (Naglieri & Das, 1997), based on the Planning, Attention, Simultaneous, and Sequential processes theory of cognition. Keith et al. examined several CFA models from both theoretical approaches. Results indicated that the hierarchical factors measured *g* at the apex and were correlated at .98. In summary, both studies found that the tests measured *g*, regardless of theoretical approach.

## 2.3 Purpose

This study compared the relationships of two tests based on different theories of ability. Research evidence indicates the MAB-II, which is based on the individual differences psychometric theory, measures *g*. The MicroCog purports to measure specific abilities based on

brain function and brain behavior relationships, *not* cognitive ability. No published studies have assessed these relationships. If these two theoretical approaches are distinct, the two tests should measure distinct constructs and should not be highly correlated. If they are highly correlated, this suggests that the two tests measure similar constructs. Further, if the MicroCog measures specific abilities, its structure should reflect these specific abilities. The MicroCog should not have a general factor that influences all of its measures. Finally, specific factors in the MicroCog should not correlate highly with measures of general ability from the MAB-II.

## **3.0 METHODS**

### **3.1 Participants**

Participants were 10,612 US Air Force pilot candidates tested on both the MAB-II and MicroCog. For those reporting demographic data, 91% were male and 84% were Caucasian. The mean age was 22 years ( $SD = 2.7$ ). All were either college graduates or enrolled in their junior or senior year of college. Test administration occurred during Medical Flight Screening (King & Flynn, 1995) prior to entry into pilot training.

### **3.2 Measures**

**3.2.1 MAB-II.** The MAB (Jackson, 1984) is a broad-based test of intellectual ability patterned after the Wechsler Adult Intelligence Scale – Revised (WAIS-R; Wechsler, 1981). It has 10 subtests that produce three summary scores: verbal IQ (VIQ), performance IQ (PIQ), and full-scale IQ (FSIQ). Previous research has demonstrated that the FSIQ score measures *g* in several age groups (Carretta et al., 1998; Kranzler, 1991; Lee, Wallbrown, & Blaha, 1990; Wallbrown, Carmin, & Barnett, 1988, 1989). The full scale IQ scores for the MAB and WAIS-R are strongly correlated ( $r = .91$ ; Conoley & Kramer, 1989; Jackson, 1984). The MAB was reviewed and restandardized in 2003 to ensure that it continued to effectively measure *g*. The result was the MAB-II, a computerized version of the test that could be individually or group administered (Jackson, 2003). Retzlaff, King, and Callister (1995) compared a paper-and-pencil version of the MAB to the computerized version and found no significant differences between them.

The MAB-II normative subtest scores have a mean of 50 and SD of 10. The normative VIQ, PIQ, and FSIQ scores have a mean of 100 and SD of 15. The test 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	Subtest	Description
Verbal IQ (VIQ)	Information (Inf)	General 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 IQ (PIQ)	Digital Symbol (Ds)	Adaptation to new set of demands; visual learning and coding, figural memory, and information processing speed
	Picture Completion (Pc)	Visual attention to detail; knowledge of common objects; perceptual and analytical skills
	Spatial (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 (Oa)	Visualization and visuo-construction skills; perceptual analytical skills needed to identify a meaningful object from left-to-right sequence

Note. From Jackson (2003).

**3.2.2 MicroCog.** The MicroCog (Powell et al., 1993) is based on the theory of brain function and brain-behavior relationships (Kaplan, 1988) and is intentionally designed to produce measures of specific abilities. This process-oriented neuropsychological approach emphasizes the distinction among cognitive domains and their underlying cortical networks. It was designed as a screening device and diagnostic tool to be used during a neuropsychological assessment of cognitive functioning.

The MicroCog is computer administered and has 18 subtests that are combined to yield 5 first-level indices: Attention/Mental Control, Reasoning/Calculation, Memory, Spatial Processing, and Reaction Time. Second and third-level indices are differentially weighted aggregates of scores from all of the first-level indices. The first-level indices each have a mean of 100 and SD of 15. These scores have been statistically adjusted by the publisher for age and education level. The test manual has well-documented internal consistency, test-retest reliability, and validity

coefficients (Powell et al., 1993). Table 2 describes the general neuropsychological aptitudes and their abbreviations for the first-level indices.

**Table 2. MicroCog Index Descriptions**

Index	Description
Spatial Processing (Spatial)	Memory for novel spatial arrangements, visuo-perceptual ability
Attention/Mental Control (Atten)	Concentration, attention span, diligence, persistence, resistance to interference
Reasoning/Calculation (Reason)	Inductive reasoning, cognitive flexibility, concept formation, basic arithmetic
Memory (Memory)	Short- and long-term memory
Reaction Time (React)	Psychomotor time between presented stimulus and response, readiness to respond, vigilance, attention

Note: From Powell et al. (1993).

### 3.3 Procedure

Participants were administered both tests prior to attending pilot training, as a routine part of Medical Flight Screening. The tests were not used for selection purposes. Rather, they were administered to capture a baseline and then archived in the event of a need to assess the individual in the future (King & Flynn, 1995). The variables chosen for analysis were the 10 MAB-II subtests and 5 MicroCog first-level indices. The MicroCog second and third-level indices cannot be analyzed simultaneously with the first-level indices because they are differentially weighted composites of the first-level indices and thus are redundant with them.

Because the participants were a highly selected group, their test scores generally have a high mean and reduced variability compared to the applicant pool. Statistics computed on these scores suffer from range restriction which causes statistical estimates to be biased with respect to population values (Bobko, Roth, & Bobko, 2001). The multivariate correction for range restriction (Lawley, 1943) was applied to the data. The MAB-II normative sample (Jackson, 1998) provided the means, SDs, and correlations used for the correction. The corrected data are superior estimates of the population values compared to the uncorrected values. This removes the bias from the selected sample estimates. The corrected values were used to conduct CFAs.

CFAs were conducted in three interdependent steps using Lisrel 8.7 (Jöreskog & Sörbom, 1996). Estimation used covariance matrices and maximum likelihood methods. First, a CFA measurement model was specified and tested for the 10 MAB-II subtests. Next, a measurement model for the 5 MicroCog scores was specified and tested. Following this, the MAB-II and MicroCog measurement models were applied to the sample that had taken both tests to determine the relationships among the factors.

Several goodness-of-fit indices were used to evaluate the measurement models: 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  $\chi^2$  was estimated, but due to its great statistical power, was not evaluated. It is presented because many of the other fit indices are based on  $\chi^2$ .

The MAB measurement model was suggested by the test developer and called for two factors corresponding to the VIQ and PIQ. See Table 1.

The MicroCog measurement model was based on an inspection of the correlations of the test scores. The positive manifold of the matrix strongly suggested a one-factor model.

The correlations of all the factors were estimated for participants who had taken both tests. Because the measurement models for the tests were evaluated independently, the goodness-of-fit estimates were not informative for the analysis having both measurement models in which the between-test factor correlations were estimated.

To estimate the proportion of  $g$  measured by each test, the first unrotated principal factor was computed independently for each test. The percent of variance associated with the first unrotated principal factor is an indication of how much the test measures  $g$  (Ree & Earles, 1991).

## 4.0 RESULTS

Table 3 provides the means, SDs, and correlations for the two tests after the data were corrected for range restriction. Observed correlations are below the diagonal and corrected correlations above it. All of the observed correlations except one were positive ( $r = -.01$ ), showing the scores were related to one another in the expected fashion. All of the corrected correlations were positive. Cohen's (1988) standards converting correlations to effect sizes ( $d$ ) were adopted for interpretation. Correlations categorized as (a) small ( $r = .10$  to  $.23$ ) have effect sizes,  $d$ , of  $.20$  to

**Table 3. Means, Standard Deviations, and Correlations of MAB-II Subtests and MicroCog First-Level Indices**

Score	Mean	SD	Inf	Com	Ari	Sim	Voc	Ds	Pc	Sp	Pa	Oa	Spatial	Atten	Reason	Memory	React
Inf	50	10	1.00	.70	.50	.62	.65	.24	.47	.31	.37	.37	.27	.40	.46	.53	.58
Com	50	10	.43	1.00	.51	.73	.63	.36	.46	.31	.41	.39	.31	.51	.49	.66	.53
Ari	50	10	.26	.28	1.00	.48	.42	.32	.36	.37	.34	.40	.35	.42	.60	.52	.23
Sim	50	10	.48	.41	.25	1.00	.61	.35	.44	.28	.40	.36	.36	.43	.51	.67	.22
Voc	50	10	.58	.44	.27	.53	1.00	.25	.39	.24	.32	.32	.25	.35	.41	.57	.07
Ds	50	10	.12	.12	.27	.18	.11	1.00	.27	.34	.36	.37	.46	.53	.45	.42	.29
Pc	50	10	.28	.22	.12	.24	.21	.22	1.00	.39	.40	.46	.35	.27	.42	.38	.31
Sp	50	10	.14	.12	.28	.14	.09	.34	.29	1.00	.36	.47	.36	.37	.47	.30	.28
Pa	50	10	.18	.14	.14	.19	.15	.25	.29	.27	1.00	.43	.32	.36	.46	.40	.27
Oa	50	10	.22	.18	.23	.21	.18	.33	.41	.40	.33	1.00	.44	.38	.52	.28	.27
Spatial	79.70	15.44	.07	.09	.17	.09	.05	.26	.17	.23	.14	.20	1.00	.46	.46	.39	.22
Atten	95.04	10.45	.16	.19	.26	.20	.18	.31	.14	.21	.13	.19	.30	1.00	.53	.53	.29
Reason	87.10	19.11	.23	.19	.37	.23	.19	.32	.22	.30	.24	.28	.24	.31	1.00	.53	.32
Memory	83.45	15.67	.29	.26	.27	.29	.30	.27	.15	.15	.16	.16	.21	.29	.31	1.00	.26
react	92.36	11.63	.02	.05	.09	.06	-.01	.20	.18	.16	.18	.15	.17	.15	.17	.08	1.00

Note: The means and SDs have been corrected for range restriction. Corrected correlations are above the diagonal; uncorrected correlations are below the diagonal.

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  $.80$  or greater.

#### 4.1 Corrected Correlations

**4.1.1 MAB-II.** All correlations among the VIQ subtests were large. The largest were between Comprehension and Similarities ( $r = .73$ ), Comprehension and Information ( $r = .70$ ), and Similarities and Vocabulary ( $r = .61$ ). All correlations among the PIQ subtests were moderate to large. The largest were between Spatial and Object Assembly ( $r = .47$ ) and Picture Completion and Object Assembly ( $r = .46$ ). The correlations between VIQ and PIQ subtests ranged from moderate (Information and Digit Symbol and Vocabulary and Spatial,  $r = .24$ ) to large (Information and Picture Completion,  $r = .47$ ; Comprehension and Picture Completion,  $r = .46$ ; and Similarities and Picture Completion,  $r = .44$ ). The strength of the correlations suggested an underlying general factor.

**4.1.2 MicroCog.** All corrected correlations were large except those involving Reaction Time. Reaction Time correlations were small with Spatial Processing ( $r = .22$ ), and moderate with all other indices. The highest correlation for Reaction Time was with Memory ( $r = .32$ ). The correlations suggested a single underlying factor affecting performance.

**4.1.3 MAB-II and MicroCog.** The corrected correlations within each test and between the tests showed positive manifold, suggesting a common factor for all measures. The smallest between test correlation was for Reaction Time and Vocabulary ( $r = .07$ ); the largest was for Memory and Similarities ( $r = .67$ ).

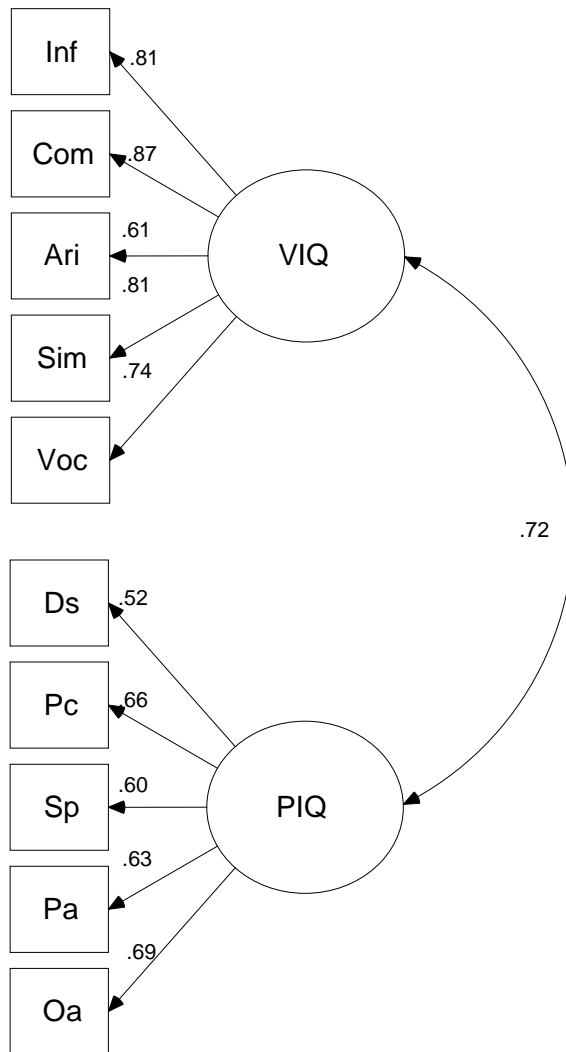
#### 4.2 Confirmatory Factor Analysis

A key purpose of this study was to examine the similarity or dissimilarity of the constructs measured by the MAB-II and the MicroCog. CFA parameter estimates converged quickly, indicating no statistical problems. As shown in Table 4, the fit of the model to the data was good

**Table 4. Goodness-of-Fit Statistics for the Two Measurement Models**

Statistic	MAB - II	MicroCog
$\chi^2$ (df)	2119.26 (34)	132.32 (5)
GFI	0.96	1.00
CFI	0.98	0.99
RMSEA	0.080 (0.076 – 0.081)	0.048 (0.041 - 0.055)
CN	281.80 ( $p < .05$ )	1,211.45 ( $p < .05$ )

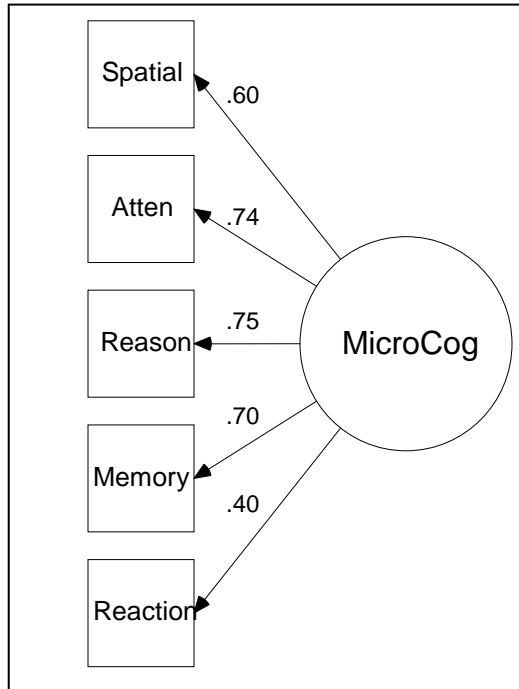
for both tests. Figure 1 shows the factor loadings and correlations between the factors for the MAB-II. Figure 2 shows the factor loadings for the MicroCog.



**Figure 1. MAB-II confirmatory factor model.**

The MAB-II model suggested by the publisher consisted of two correlated factors interpreted as Verbal (FVIQ) and Performance (FPIQ). The 5 MicroCog indices contributed to a single factor.

Figure 3 shows the joint factor structure and correlations among the factors. The correlations were high and positive, affirming the commonality of measured constructs. Figure 4 shows the hierarchical factor structure of the first-order factors of the MAB-II and the MicroCog, with *g* at its apex.

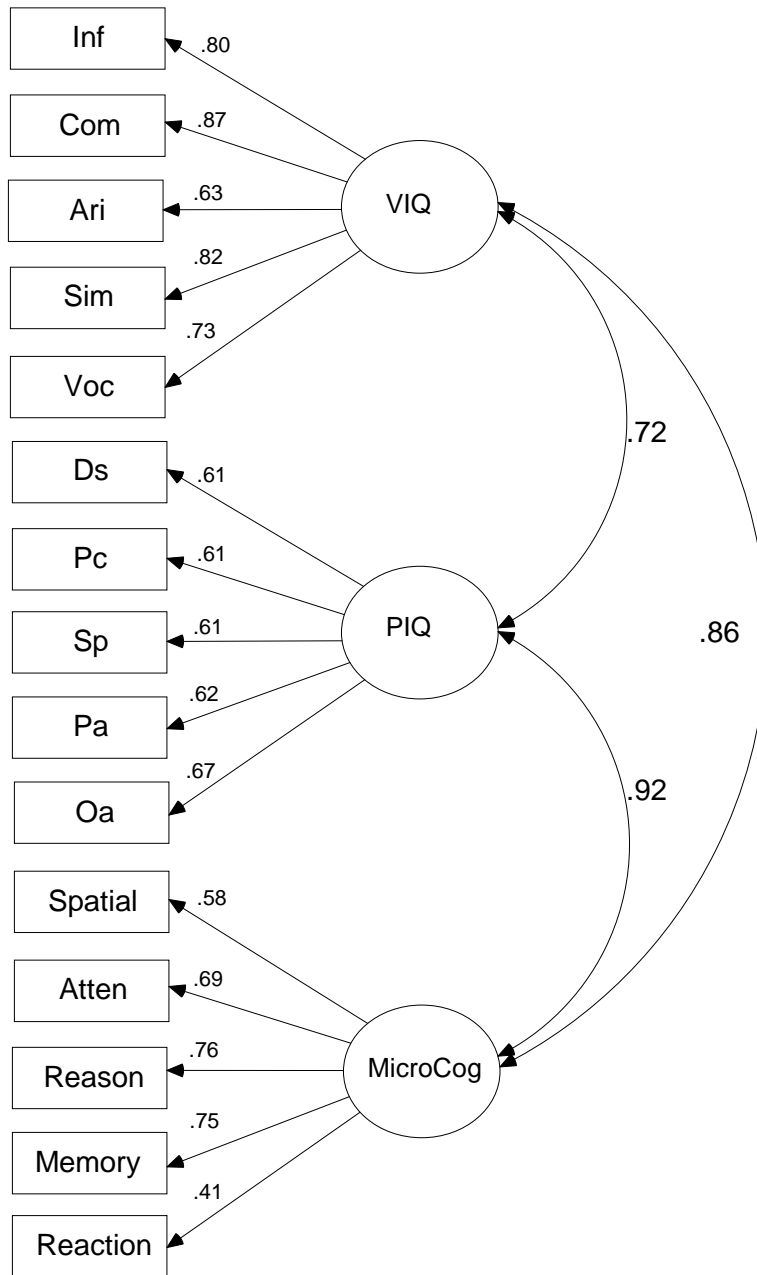


**Figure 2. MicroCog confirmatory factor model.**

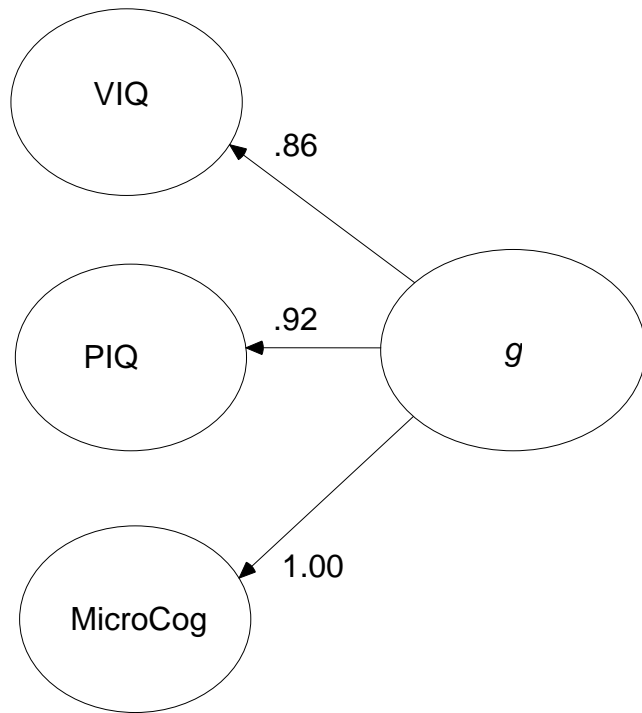
Principal factors analyses were conducted to determine the common variance in each factor attributable to *g* (Ree & Earles, 1991). The common variance in the MAB-II FVIQ and FPIQ factors were 52% and 45%, respectively. The proportion of the common variance in the scores due to *g* was 48%, for all 10 MAB-II subtests. The MicroCog reports 5 first-level indices. The proportions of common variance for the indices were; Reasoning, 42%, Attention, 41%, Memory, 38%, Spatial, 29%, and Reaction Time 13%. Fully 52% of the common variance for all five MicroCog indices was due to *g*.

## 5.0 DISCUSSION

Over the last century, theories of ability progressed from two factors (*g* and *s*) to multiple factors and, finally, to hierarchical factors that influenced lower-order factors. The current study examined the similarity of two tests designed using two different theories of ability. These two theories were developed from very different assumptions about the nature of abilities. One test was based on individual differences psychometric theory while the other was based on a specific abilities theory grounded in brain-behavior relationships. Correlational results indicated that positive manifold occurred for both tests for both the observed and corrected correlations. Positive manifold can be used to guide analyses because it is indicative of a general factor such as *g*.



**Figure 3. Joint factor structure of the MAB-II and MicroCog.**



**Figure 4. Apex of the hierarchical model for the MAB-II and MicroCog.**

CFAs were conducted to determine whether the theoretical bases underlying the construction of the two tests produced similar or different constructs. The measurement models for each test fit the data well.

Examination of the three factors derived from the MAB-II and MicroCog CFAs showed strong correlations between the single MicroCog factor and the MAB PIQ ( $r = .92$ ) and VIQ ( $r = .86$ ) factors. These correlations among the first-order factors suggested substantial overlap of constructs measured and that a higher-order factor influenced all of the lower-order factors.

The proportion of variance due to *g* was consistent with previous studies (Carretta, Callister, & King, 1998; Carretta & Ree, 1996; Ree & Carretta, 1994), in which the common variance (*g*) accounted for about half or more of the variability in the scores. The difference in the proportion of variance due to *g* between the two tests was very small, indicating they each measure *g* in about the same amount. The similarity of the constructs measured was further reinforced when the hierarchical *g* factor was included in the joint model. The *g* loading of each of the three lower-order factors approached or was equal to 1. These results indicated that ability tests constructed using two different theoretical approaches produced very similar results.

Other studies have compared tests developed from an individual differences psychometric approach to tests based on other approaches such as cognitive components (Stauffer et al., 1996) and individual differences *processes* theory (Keith et al., 2001). The current study compared a test developed from an individual differences psychometric theory to one based on a theory of brain-behavior relationships. The theories of cognitive components, processes, and brain-behavior relationships are all theories of specific abilities. Despite different theoretical approaches to test development, the resultant tests measured a well-known construct, *g*.

One explanation for the consistency of these results can be found in what Spearman (1923) called the “indifference of the indicator” or “indifference of the fundamentals.” This means that test *content* is not fundamental to the measurement of *g*. It is the vehicle that allows for the expression of relationships and differences measured in cognitive ability tests. Indeed, cognitive ability can and has been measured with verbal, quantitative, and spatial test items of widely varying content. Tests of reasoning such as Raven’s Matrices (Raven, 1939) have no verbal or quantitative content, and little spatial content yet it measures *g*. Chronometric measures (i.e., cognitive speed) requiring no verbal, quantitative, or spatial content also have been shown to measure *g* (Jensen, 2006). Hence, indifference of the indicator has been firmly established in empirical studies and was confirmed in the current study.

These findings add to both test development theory and the use of cognitive tests. The accumulation of research evidence shows that despite attempts to develop measures of specific ability, *g* accounts for about 50% of the variance in cognitive ability tests. Variance due to specific abilities may be present in much smaller quantities. Even tests designed to measure specific abilities have been shown to measure *g*.

In a recent examination of the empirical evidence on general and specific cognitive abilities, Ree et al. (2013) called for researchers intent on developing new theories and measures of specific abilities to provide empirical evidence that: (1) the specific abilities exist independent of well-known constructs such as *g*; and (2) the specific abilities are not combinations of specific variance and variance due to *g*. This research contributes to the theoretical understanding of cognitive ability tests and their structure by demonstrating empirically that the MicroCog, designed to measure specific cognitive abilities, measures *g* in about the same proportions as a test developed to measure mostly *g*. This is consistent with Spearman's (1923) “indifference of the indicator” concept.

The implications for testing practice are equally compelling. Evidence suggests that measures of *g* used for selection provide the strongest prediction of occupational and educational criteria (Hunter, 1986). Specific abilities add little to *g* for the prediction of training or job performance (Jensen, 1980, 1998; McNemar, 1964; Ree & Earles, 1991, 1992; Ree et al., 1994; Schmidt & Hunter, 1998). For practical purposes, an important distinction must be made between specific

*abilities* and specific *knowledge*. Specific ability should not be confused with specific knowledge, which can be incremental to the predictive validity of *g*. General cognitive ability is required to acquire specific knowledge about a particular field and *g* is a precursor to the acquisition of specific knowledge (Hunter, 1986; Ree et al., 1995). Resources spent developing measures of specific *ability* for purposes of selection might be better allocated to the development of measures of specific *knowledge*.

## 6.0 REFERENCES

- Bobko, P., Roth, P. L., & Bobko, C. (2001). Correcting the effect size of  $d$  for range restriction and unreliability. *Organizational Research Methods*, 4, 46-61.
- 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., Callister, J. D., & King, R. E. (1998). A comparison of two U. S. Air Force pilot aptitude tests. *Aviation, Space, and Environmental Medicine*, 69, 931-935.
- Carretta, T. R., Retzlaff, P. D., Callister, J. D., & King, R. E. (1998). A comparison of two U.S. Air Force pilot tests. *Aviation, Space, and Environmental Medicine*, 69, 931-935.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.
- Galton, F. (1869). *Hereditary genius: An inquiry into its laws and consequences*. London: Macmillan.
- Hull, C. L. (1928). *Aptitude testing*. Yonkers-on-the-Hudson, NY: World Book.
- Hunter, J. E. (1986). Cognitive ability, cognitive aptitude, job knowledge, and job performance. *Journal of Vocational Behavior*, 29, 340-362.
- Hunter, J. E., & Hunter, R. F. (1984). Validity and utility of alternative predictors of job performance. *Psychological Bulletin*, 96, 72-98.
- 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.
- Jensen, A. R. (1980). *Bias in mental testing*. NY: Free Press.
- Jensen, A. R. (1998). *The g factor: The science of mental ability*. Westport, CT: Praeger.
- Jensen, A. R. (2006). *Clocking the mind: Mental chronometry and individual differences*. Oxford, UK: Elsevier.
- Jöreskog, K., & Sörbom, D. (1996). *LISREL 8: User's reference guide*. Chicago: Scientific Software International, Inc.
- Kaplan, E. (1988). A process approach to neuropsychological assessment. In T. Boll & B.K. Bryant (Eds.), *Clinical neuropsychology and brain function: Research, measurement and practice*. Washington, DC: American Psychological Association.
- Keith, T. Z., Kranzler, J. H., & Flanagan, D. P. (2001). What does the Cognitive Assessment System (CAS) measure? Joint confirmatory factor analysis of the CAS and the Woodcock-Johnson Tests of Cognitive Ability (3<sup>rd</sup> ed.). *School of Psychology Review*, 30, 89-119.

- King, R. E., & 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. (1993). Aptitude testing inspired by information processing: A test of the four-sources model. *Journal of General Psychology*, *120*, 375-405.
- Kyllonen, P. C. (1994). CAM: A theoretical framework for cognitive abilities measurement. In D. Detterman (Ed.), *Current topics in human intelligence: Volume 4: Theories of intelligence* (pp. 307-359). Norwood, NJ: Ablex.
- Lance, C. E., & Vandenberg, R. J. (Eds.) (2009). *Statistical and methodological myths and urban legends: Doctrine, verity and fable in the organizational and social sciences*. NY: Routledge.
- Lawley, D. N. (1944). A note on Karl Pearson’s selection formulae. *Proceedings of the Royal Society of Edinburgh. Section A. Mathematical and Physical Sciences*, *62*, 28-30.
- Lee, M. S., Wallbrown, F. H., & Blaha, J. (1990). Note on the construct validity of the Multidimensional Aptitude Battery. *Psychological Reports*, *67*, 1219-1222.
- McNemar, Q. (1964). Lost: Our intelligence? Why? *American Psychologist*, *19*, 871-882.
- Murphy, R. R. (2009). Content validation is useful for many things, but validity isn’t one of them. *Industrial and Organizational Psychology*, *2*, 453 – 464.
- Naglieri, J. A., & Das, J. P. (1997). *Cognitive Assessment System*. Itasca, IL: Riverside.
- Powell, D. H., Kaplan, E. F., Whitla, D., Weintraub, S., Caitlin, R., & Funkenstein, H. H. (1993). *MicroCog: Assessment of cognitive functioning version 2.1*. San Antonio, TX: The Psychological Corporation.
- Raven, J. C. (1939). *Progressive matrices: A perceptual test of intelligence*. London: H. K. Lewis.
- Ree, M. J., & Carretta, T. R. (1994). Factor analysis of the ASVAB: Confirming a Vernon-like structure. *Educational and Psychological Measurement*, *54*, 459-463.
- Ree, M. J., Carretta, T. R., & Teachout, M. S. (1995). Role of ability and prior job knowledge in complex training performance. *Journal of Applied Psychology*, *80*, 721-730.
- Ree, M. J., & Earles, J. A. (1991). The stability of *g* across different methods of estimation. *Intelligence*, *15*, 271-278.
- Ree, M. J., & Earles, J. A. (1992). Intelligence is the best predictor of job performance. *Current Directions in Psychological Science*, *1*, 86-89.
- Ree, M. J., Earles, J. A., & Teachout, M. S. (1994). Predicting job performance: Not much more than *g*. *Journal of Applied Psychology*, *79*, 518-524.
- Ree, M. J., Teachout, M. S., & Carretta, T. R. (2013). Against all evidence: General and specific ability in human resource management. In D. J. Svantek, & K. T. Mahoney

- (Eds), *Received wisdom: Kernels of truth and boundary conditions in organizational studies* (pp. 181-199). Information Age Publishing.
- Retzlaff, P. D., King, R. E., & Callister, J. D. (1995). *USAF pilot training completion and retention: A ten year follow-up on psychological testing*, AL/AO-TR-1995-0124. Armstrong Laboratory, Aerospace Medicine Directorate, Brooks AFB, TX.
- Sanders, S., McIntosh, D. E., Dunham, M., Rothlisberg, B. A., & Finch, H. (2007). Joint confirmatory factor analysis of the differential ability scales and the Woodcock-Johnson Tests of Cognitive Abilities (3<sup>rd</sup> ed.). *Psychology in the Schools*, *44*, 119-138.
- Schmidt, F. L., & Hunter, J. E. (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological Bulletin*, *124*, 262-274.
- Spearman, C. (1904). "General intelligence," objectively determined and measured. *American Journal of Psychology*, *15*, 201-293.
- Spearman, C. (1923). *The nature of "intelligence" and the principles of cognition*. London: Macmillan.
- Spearman, C. (1927). *The abilities of man: Their nature and measurement*. NY: Macmillan.
- Sperl, T. C., Ree, M. J., & Steuck, K. W. (1992). Armed Services Vocational Aptitude Battery and Air Force Officer Qualifying Test: Analyses of common attributes. *Military Psychology*, *4*, 175-188.
- Stauffer, J. M., Ree, M. J., & Carretta, T. R. (1996). Cognitive-components tests are not much more than g: An extension of Kyllonen's analyses. *Journal of General Psychology*, *123*, 193-205.
- Thurstone, L. L. (1938). *Primary mental abilities*. Chicago, IL: University of Chicago Press.
- Vernon, P. E. (1950). *The structure of human abilities*. NY: Wiley.
- Vernon, P. E. (1969). *Intelligence and cultural environment*. London: Methuen.
- 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. (1980). *Wechsler Adult Intelligence Scale – Revised (WAIS-R)*. San Antonio, TX: Harcourt Assessment.
- Woodcock, R. W., McGrew, K. S., & Mather, N. (2001). *Woodcock-Johnson III*. Itasca, IL: Riverside.

## LIST OF ACRONYMS, ABBREVIATIONS, AND SYMBOLS

AFOQT	Air Force Officer Qualifying Test
ASVAB	Armed Services Vocational Aptitude Battery
CN	Critical N
<i>d</i>	Standardized mean score difference
<i>g</i>	General cognitive ability
CFA	Confirmatory Factor Analysis
CFI	Comparative Fit Index
GFI	Goodness of Fit Index
MAB	Multidimensional Aptitude Battery
MAB Ari	MAB Arithmetic subtest
MAB Com	MSB Comprehension subtest
MAB Ds	MAB Digit Symbol subtest
MAB FSIQ	MAB Full-Scale Intelligence Quotient
MAB Inf	MAB Information subtest
MAB Oa	MAB Object Assembly subtest
MAB Pa	MAB Picture Arrangement subtest
MAB Pc	MAB Picture Completion subtest
MAB PIQ	MAB Performance Intelligence Quotient
MAB Sim	MAB Similarities subtest
MAB Sp	MAB Spatial subtest
MAB VIQ	MAB Verbal Intelligence Quotient
MAB Voc	MAB Vocabulary subtest
MicroCog Atten	MicroCog Attention/Mental Control index
MicroCog Memory	MocoCog Memory index
MicroCog React	MicroCog Reaction Time index
MicroCog Reason	MicroCog Reasoning/Calculation index
MicroCog Spatial	MicroCog Spatial Processing index
<i>r</i>	Correlation coefficient
<i>s</i>	Specific abilities
RMSEA	Root Mean Square Error of Approximation
WAIS-R	Wechsler Adult Intelligence Scale-Revised