Wright State University

CORE Scholar

International Symposium on Aviation Psychology - 2013

International Symposium on Aviation Psychology

2013

# Pilot Cognitive Functioning and Training Outcomes

Thomas R. Carretta

Raymond E. King

Paul D. Retzlaff

Erica Barto

Malcom James Ree

*See next page for additional authors*

Follow this and additional works at: https://corescholar.libraries.wright.edu/isap_2013

Part of the Other Psychiatry and Psychology Commons

## Repository Citation

Carretta, T. R., King, R. E., Retzlaff, P. D., Barto, E., Ree, M. J., & Teachout, M. S. (2013). Pilot Cognitive Functioning and Training Outcomes. *17th International Symposium on Aviation Psychology*, 591-596. https://corescholar.libraries.wright.edu/isap_2013/16

This Article is brought to you for free and open access by the International Symposium on Aviation Psychology at CORE Scholar. It has been accepted for inclusion in International Symposium on Aviation Psychology - 2013 by an authorized administrator of CORE Scholar. For more information, please contact library-corescholar@wright.edu.

## Authors

Thomas R. Carretta, Raymond E. King, Paul D. Retzlaff, Erica Barto, Malcom James Ree, and Mark S. Teachout

This article is available at CORE Scholar: https://corescholar.libraries.wright.edu/isap_2013/16

# PILOT COGNITIVE FUNCTIONING AND TRAINING OUTCOMES

Thomas R. Carretta[1] Raymond E. King[2] Paul D. Retzlaff[3]
Erica Barto[4] Malcolm James Ree[4] Mark S. Teachout[5]

The predictive validity of scores from two cognitive functioning tests, the Multidimensional Aptitude Battery and MicroCog, was examined for initial pilot training outcomes. In addition to training completion, academic grades, daily flying grades, check ride grades, and class rank were available for graduates. Mean score comparisons and correlations in samples of between 5,582 and 12,924 trainees across the two tests showed small, but statistically significant, relationships with training performance. The results pointed to general cognitive ability as the main predictor of training performance. Comparisons with results from studies involving US Air Force pilot aptitude tests showed lower validities for these cognitive functioning tests. This finding likely occurred because the pilot aptitude tests measure additional factors (e.g., aviation knowledge/experience, psychomotor) that are predictive of training success, but not measured by these cognitive functioning tests, which were designed primarily to be used for clinical assessment.

Measures of cognitive ability have been a mainstay in military pilot aptitude batteries since WWI (Carretta & Ree, 2003). Although the specific content and administration mode vary, cognitive ability has shown a consistent relation with pilot performance (Carretta & Ree, 2003; Hunter & Burke, 1995; Martinussen, 1996). More recently, Paullin, Katz, Bruskiewicz, Houston, and Damos (2006) conducted a comprehensive review of aviation testing and selection for the US Army that included both cognitive and personality tests. They recommended the US Army follow the lead of the US Navy and US Air Force in their use of selection tests and that they focus on measures of intelligence, cognitive ability, and information processing. Howse and Damos (2011) updated that work with a comprehensive, 275-page annotated bibliography published through the Air Force Personnel Center. These reviews and other studies (Olea & Ree, 1995; Ree & Carretta, 1996; Zierke, 2012) have shown intelligence and cognitive ability to be crucial to pilot training performance. Additional predictors include aviation knowledge/experience, psychomotor ability, and, perhaps, personality (Carretta & Ree, 2003; Hunter & Burke, 1995; Martinussen, 1996).

## US Air Force (USAF) Pilot Trainee Selection

All USAF pilot training applicants must pass a rigorous Class I flight physical (USAF, 2011) to be eligible for selection. Medically qualified applicants are evaluated for training suitability on measures of aptitude and officership (Weeks & Zelenski, 1998). USAF Academy (USAFA) cadets are evaluated by faculty and staff, who consider academic, physical, and military performance. Applicants commissioned through the Reserve Officer Training Corps (ROTC) or Officer Training School (OTS) are administered the Air Force Officer Qualifying Test (AFOQT; Drasgow, Nye, Carretta, & Ree, 2010) and Test of Basic Aviation Skills (TBAS; Carretta, 2005). The AFOQT Pilot composite, several TBAS subtest scores, and a measure of flying experience are combined n a regression-weighted equation to create a measure of pilot training aptitude called the Pilot Candidate Selection Method (PCSM; Carretta, 2011). For ROTC, medically qualified applicants are ranked on an Order of Merit score based on the PCSM score, field training, physical fitness, college Grade Point Average (GPA), and commander's ranking. OTS pilot candidate selection uses the "whole person" concept, where applicants receive points for experience/leadership, education/aptitude, and potential/adaptability. All of these selection procedures emphasize high intelligence, whether it involves acceptance into the USAFA, a high GPA, a high AFOQT score, or the impression a candidate makes on a selection board.

**Air Force Officer Qualifying Test.** The current AFOQT (Form S) has 11 cognitive subtests used to create five composites: Verbal (V), Quantitative (Q), Academic Aptitude (AA), Pilot (P), and Combat Systems Officer (CSO). The V and Q composites are used to qualify civilians and prior-enlisted USAF personnel for officer commissioning through the OTS and ROTC programs. The P and CSO composites are used to qualify applicants who pass other educational, aptitude, and physical requirements for aircrew training. The AFOQT has a hierarchical factor structure and measures general cognitive ability (*g*) and the lower order factors of verbal, math, spatial, aircrew interest/ aptitude, and perceptual speed (Drasgow et al., 2010). It has been validated for officer training (Roberts & Skinner, 1996), aircrew training (Carretta, 2008, 2013; Carretta & Ree, 1995, 2003; Olea & Ree, 1994), and for several non-

[1] Air Force Research Laboratory, Wright-Patterson AFB, OH; [2]USAF School of Aerospace Medicine, Wright-Patterson AFB, OH;; [3]Eagle Applied Sciences, San Antonio, TX; [4]Operational Technologies Corp, San Antonio, TX; [5]University of the Incarnate Word, San Antonio, TX

aircrew jobs (Carretta, 2010). Its predictive validity for pilot training comes from its measurement of cognitive ability and pilot job knowledge (Carretta & Ree, 2003; Olea & Ree, 1994; Ree & Carretta, 1996).

Despite its effectiveness for measuring cognitive ability and its utility for officer and aircrew training qualification, AFOQT scores are not easily interpretable in ideographic assessment. USAF clinicians prefer tests such as the Multidimensional Aptitude Battery (MAB; Jackson, 1984, 1988) to the AFOQT in such assessment due to its similarity to tests such as the Wechsler Adult Intelligence Scale - Revised (WAIS-R). Clinicians find the MAB relatively easy to use to make pre- and post-incident comparisons due to its similarity to the WAIS-R.

**Cognitive Testing.** While accession procedures focus on intelligence, so does much of the ideographic assessment of pilot candidates. The USAF Medical Flight Screening (MFS) program screens pilot candidates prior to Specialized Undergraduate Pilot Training (SUPT). In addition to ophthalmic and cardiac diagnostic procedures, several cognitive and personality tests are administered (King, Barto, Ree, & Teachout, 2011; King, Barto, Ree, Teachout, & Retzlaff, 2011). The primary purpose of the cognitive tests is to archive cognitive functioning data for future use. The intent is to develop a registry against which future testing might be compared. The psychological portion of MFS includes both traditional measures of intelligence and computerized cognitive tasks.

As the primary purpose of the psychological testing is to enable potential ideographic assessment, there has been little emphasis on training outcomes. To date, the MFS cognitive tests have not been validated against pilot training outcomes. Boyd, Patterson, and Thompson (2005), however, evaluated some of the tests against aircraft type later flown. Usually, fighter/bomber aircraft advanced training assignments are offered to those highest in class rank in primary jet training. Class rank accounts for much of the variance in advanced training assignments although other factors (e.g., number of fighter/non-fighter training slots, student preferences, and Guard/Reserve pilots flying what their squadrons fly) also affect advanced training assignments. Boyd et al. (2005) compared one of the MFS cognitive tests, the MAB (Jackson, 1998), and one of the personality tests, the NEO PI-R (Costa & McCrae, 1985), to airframe assignment (fighter, bomber, and airlift/tanker). Small (Cohen, 1988), but statistically significant differences were observed between the groups, with the mean IQs for fighter pilots 2 to 3 points (about .13 to .20 SDs) higher than for airlift/tanker pilots. Using the means, SDs, and sample sizes reported by Boyd et al., we converted the differences to a correlation statistic (Lipsey & Wilson, 2000). The mean difference in verbal, performance, and full scale IQ between those assigned to fighters vs. airlift/tankers were equivalent to correlations of .14, .15, and .18. These results suggest intelligence has a modest relationship with advanced training assignment.

### Purpose

The purpose of the current study was to determine the extent to which two tests used to assess cognitive functioning and typically the domain of clinical assessment, the MAB and MicroCog, predict USAF initial (T-6) pilot training outcomes. Separate validation studies were done in order to maximize the sample sizes for each test.

## Methods

### Participants

Participants were USAF personnel selected for SUPT that had tested on the MAB, MicroCog, or both. The sample sizes were 12,924 for the MAB and 5,582 for the MicroCog. All participants were college graduates or near completion of college. Sample demographics were similar for the two studies. Of those reporting demographic data, 91% were male. They had a mean age of 23 years, and 99% were 30 years of age or less. Eighty-four percent reported they were white. Test administration either occurred at the USAFA or at the USAF School of Aerospace Medicine (USAFSAM) prior to entry into SUPT. The T-6 completion rate was about 89.5% for both samples.

### Measures

**Multidimensional Aptitude Battery.** The MAB (Jackson, 1984, 1998) is a broad-based test of cognitive ability patterned after the WAIS–R (Wechsler, 1981). The full-scale IQ scores for the MAB and WAS-R are highly correlated ($r = .91$; Conoley & Kramer, 1989; Jackson, 1984). The MAB can be individually or group administered and requires less than 1.5 hours. Its 10 subtests produce three scores: full-scale IQ (FSIQ), verbal IQ (VIQ), and performance IQ (PIQ). The IQ scores have a mean of 100 and a SD of 15 in the general population. Test-retest reliability for the IQ scores ranges from .94 to .98 (Jackson, 1998) for a retest interval averaging 45 days. The FSIQ score has been shown to measure *g* in several age groups (Wallbrown, Carmin, & Bartlett, 1988, 1989). Chappelle, McDonald, Thompson, McMillan, and Marley (2010) examined the MAB for USAF gunship sensor operators and found no mean differences between training graduates and eliminees.

**MicroCog.** The MicroCog (Powell, Kaplan, Whitla, Weintraub, Caitlin, & Funkenstein, 1993) is a computer-administered cognitive functioning test that assesses a range of cognitive behaviors such as reaction time and memory. The primarily purpose of the test was to assess clinical pathology in patients. While the MAB is a classic IQ test, the MicroCog comes more from a clinical neuropsychological perspective (Vanderploeg, 2000).

The MicroCog has 18 subtests combined to create nine index scores. The indices take two forms, domain-based and higher-order summary scores. The five domains are Attention/Mental Control, Memory, Reasoning/Calculation, Spatial Processing, and Reaction Time. The four higher-order summary scores are Information Processing Speed (IPS), Information Processing Accuracy (IPA), General Cognitive Functioning (GCF), and General Cognitive Proficiency (GCP). IPS and IPA reflect a potential two-factor structure of the subtests. GCF and GCP are purported to represent general cognitive ability, where GCF is a function of the two Information Processing scores and GCP is a summation of the Proficiency scores of all the subtests (Powell et al., 1993). The Information Processing and General Cognitive indices generally correlate with the WAIS-R in the .50s.

Chappelle, Ree, Barto, Teachout, and Thompson (2010) compared the MAB and MicroCog using structural equation models. They concluded that both tests have a factor representing $g$. The MicroCog only produced one factor, suggesting there is less specificity to the scores than may be desired by clinicians or researchers. Inasmuch as the MicroCog appears to measure only one factor, and due to space limitations, we focused on the four higher-order summary scores in our analyses.

**Pilot training criteria.** Several SUPT initial jet training (T-6) performance criteria were examined. There were three dichotomous training completion scores for graduates and eliminees: graduation/elimination, graduation/ flying training deficiency (FTD) elimination, and graduation/drop on request (DOR) elimination. Several additional criteria were available only for graduates: academic grades, daily flying grades, check flight grades, and class rank. Class rank is a weighted average of academic, daily flying, and check flight grades. In computing class rank, flying grades get more weight than academic grades and check flight grades get more weight than do daily flying grades.

### Analyses

Analyses began with examination of the means and SDs for the cognitive test scores. Univariate statistics were used to determine the relations of the cognitive test scores to the training performance criteria. All statistical analyses used a .05 Type I error rate and one-tailed tests. Next, the observed correlations between the test scores and training criteria were corrected for range restriction using the multivariate method (Lawley, 1943). The MAB and MicroCog scores could not be corrected to the same reference group as the participants lacked a common selection test (e.g., AFOQT). As a result, the data for each test were corrected to the respective normative group. After correction for range restriction, the correlations involving the test scores and training completion criteria were corrected for dichotomization (Cohen, 1983). The correlations involving the test scores and pilot training grades were corrected for unreliability (Hunter & Schmidt, 2004) of the training criteria ($r_c = \frac{r_{xy}}{\sqrt{r_{yy}}}$). The reliability of the training grades was estimated at .80 based on results from similar studies that examined academic grades (Kuncel, Hazlett, & Ones, 2001, 2004). The correlations corrected for range restriction and reliability of the training grades provide a theoretical estimate of the predictiveness of the test scores when a perfectly reliable criterion is available.

## Results

### Study 1: Multidimensional Aptitude Battery

**Graduation vs. elimination.** The overall graduation rate was 89.6% (11,579/12,924). When only graduates and either FTD or DOR eliminees were included the graduation rates were 95.4% (11,579/12,138) and 95.9% (11,579/12,079) respectively.

The MAB IQ scores were severely range restricted compared to the normative values where the means and SDs are 100 and 15. The IQ scores for graduates and each of the eliminee groups were high at about 120 (about 1.33 SDs above the normative mean) and the variances of the scores were much less than the normative values. For the FSIQ score the variance for the trainees was about 18% of the normative value. All mean score differences between graduates and eliminees favored graduates, but were small (i.e., about 2 points for the IQ scores). Despite this, all mean score comparisons were statistically significant. Larger differences occurred for graduates vs. FTD eliminees than for graduates vs. DOR eliminees. This result may be because DOR elimination may occur for reasons not related to ability (e.g., motivation). Examination of the observed correlations indicated all effect sizes were small (< .10; Cohen, 1989). While very large samples ensure sufficient statistical power, very small differences will be statistically significant yet may offer little practical predictive power. Low point-biserial correlations for the IQ scores reinforce the small mean score differences. A .083 correlation was observed between the FSIQ score and the graduation/elimination criterion. It should be noted that the training eliminees included medical and self-elimination losses, so the group distinctions in this analysis are not as clear as desired.

Table 1 summarizes the observed and corrected correlations. As expected, the correlations increased in magnitude after correction. For example, the correlation between the FSIQ score and graduation/elimination was .083 in the observed data, increased to .192 after correction for range restriction, and to .323 after correction for both range restriction and dichotomization of the criterion. Similar trends occurred for the other MAB scores and criteria.

Table 1. *MAB: Observed and Corrected Correlations with Training Completion*

| Score | Graduation/All Eliminees | | | Graduation/FTD Eliminees | | | Graduation/DOR Eliminees | | |
|---|---|---|---|---|---|---|---|---|---|
| | r | $r_c$ | $r_{fc}$ | r | $r_c$ | $r_{fc}$ | r | $r_c$ | $r_{fc}$ |
| FSIQ | $.083^b$ | .208 | .295 | $.072^b$ | .158 | .264 | $.039^b$ | .106 | .202 |
| VIQ | $.057^b$ | .185 | .262 | $.042^b$ | .115 | .192 | $.022^a$ | .101 | .192 |
| PIQ | $.079^b$ | .199 | .282 | $.074^b$ | .165 | .275 | $.042^b$ | .104 | .198 |

*Notes*. The column headings indicate observed correlations (r), correlations corrected for range restriction ($r_c$), and correlations corrected for both range restriction and dichotomization of the criterion ($r_{fc}$). Statistical significance was tested only for the observed correlations. $^a p < .05$; $^b p < .01$

**Training grades.** As shown in Table 2, all observed correlations between the test scores and training grades were statistically significant. FSIQ had the strongest observed correlation for all of the training grades. The strongest correlations for the IQ scores occurred for academic grades (e.g., FSIQ; r = .233). The FSIQ correlation with class rank, which is a weighted average of the academic and flying training grades, was .157. All correlations increased in magnitude after correction for range restriction and again after correction for both range restriction and reliability of the criteria. After correction for both range restriction and reliability of the criteria, FSIQ was correlated .551 with academic grades, .316 with daily flying grades, .282 with check flight grades, and .374 with class rank.

Table 2. *MAB: Observed and Corrected Correlations with Training Grades*

| Score | Academic Grades | | | Daily Flying Grades | | | Check Flight Grades | | | Class Rank | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | r | $r_c$ | $r_{fc}$ | r | $r_c$ | $r_{fc}$ | r | $r_c$ | $r_{fc}$ | r | $r_c$ | $r_{fc}$ |
| FSIQ | $.233^b$ | .505 | .564 | $.124^b$ | .257 | .287 | $.110^b$ | .235 | .262 | $.157^b$ | .325 | .363 |
| VIQ | $.224^b$ | .497 | .555 | $.084^b$ | .229 | .256 | $.083^b$ | .217 | .242 | $.123^b$ | .308 | .344 |
| PIQ | $.164^b$ | .428 | .478 | $.120^b$ | .252 | .281 | $.099^b$ | .222 | .248 | $.138^b$ | .304 | .339 |

*Notes*. The column headings indicate observed correlations (r), correlations corrected for range restriction ($r_c$), and correlations corrected for both range restriction and reliability of the criterion ($r_{fc}$). Statistical significance was tested only for the observed correlations. N = 11,579; $^a p < .05$; $^b p < .01$

**Study 2: MicroCog**

**Graduation vs. elimination.** The overall graduation rate was 89.4% (4,992/5,582). When only graduates and either FTD or DOR eliminees were included the graduation rates were 93.5% (4,992/5,238) and 96.1% (4,992/5,194) respectively.

Although the MicroCog scores were affected by range restriction, the amount of restriction was less than that for the MAB. Both tests have means and SDs of 100 and 15 in their respective normative samples. However, whereas the average means and SDs for the MAB IQ scores were about 120 and 6.4 for pilot trainees, the average means and SDs for the MicroCog scores were about 104 and 11. The variances of the MAB and MicroCog scores were respectively about 18% and 54% that for their respective normative populations. The difference in amount of restriction on the two tests was likely due to differences in the composition of the normative groups. MicroCog population norms are based on scores corrected for age and education level. All mean score comparisons between graduates and eliminees favored graduates and were statistically significant for the analyses involving all eliminees

Table 3. *MicroCog: Observed and Corrected Correlations with Training Completion*

| Score | Graduation/All Eliminees | | | Graduation/FTD Eliminees | | | Graduation/DOR Eliminees | | |
|---|---|---|---|---|---|---|---|---|---|
| | r | $r_c$ | $r_{fc}$ | r | $r_c$ | $r_{fc}$ | r | $r_c$ | $r_{fc}$ |
| IPS | $.068^b$ | .125 | .201 | $.077^b$ | .129 | .252 | $.024^a$ | .050 | .112 |
| IPA | $.053^b$ | .137 | .220 | $.063^b$ | .147 | .287 | .015 | .050 | .112 |
| GCF | $.083^b$ | .168 | .270 | $.097^b$ | .178 | .347 | $.026^a$ | .063 | .141 |
| GCP | $.091^b$ | .165 | .266 | $.102^b$ | .171 | .334 | $.031^a$ | .065 | .145 |

*Notes*. The column headings indicate observed correlations (r), correlations corrected for range restriction ($r_c$), and correlations corrected for both range restriction and dichotomization of the criterion ($r_{fc}$). Statistical significance was tested only for the observed correlations. $^a p < .05$; $^b p < .01$

and FTD eliminees. Smaller mean score differences occurred between graduates and DOR eliminees. As with the MAB, all point-biserial correlations effect sizes were small (< .10; Cohen, 1989).

Table 3 summarizes the observed and corrected correlations. Even after correction for both range restriction and dichotomization, only eight of 12 correlations were above .20; only two were above .30.

**Training grades.** As with the MAB, all observed correlations between the test scores and training criteria were statistically significant. See Table 4. GCF and GCP demonstrated the highest predictive validities averaged across the training criteria. The strongest correlations for three of the four MicroCog scores occurred for academic grades. After correction for both range restriction and reliability of the criteria, GCF was correlated .341 with academic grades, .285 with daily flying grades, .251 with check flight grades, and .333 with class rank.

Table 4. *MicroCog: Observed and Corrected Correlations with Training Grades*

| Score | Academic Grades | | | Daily Flying Grades | | | Check Flight Grades | | | Class Rank | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $r$ | $r_c$ | $r_{fc}$ | $r$ | $r_c$ | $r_{fc}$ | $r$ | $r_c$ | $r_{fc}$ | $r$ | $r_c$ | $r_{fc}$ |
| IPS | .075$^b$ | .164 | .183 | .137$^b$ | .208 | .232 | .102$^b$ | .164 | .183 | .139$^b$ | .219 | .244 |
| IPA | .220$^b$ | .314 | .351 | .093$^b$ | .191 | .213 | .097$^b$ | .182 | .203 | .138$^b$ | .246 | .275 |
| GCF | .206$^b$ | .305 | .341 | .165$^b$ | .255 | .285 | .149$^b$ | .225 | .251 | .202$^b$ | .298 | .333 |
| GCP | .204$^b$ | .299 | .334 | .170$^b$ | .249 | .278 | .148$^b$ | .219 | .244 | .201$^b$ | .289 | .323 |

*Notes*. The column headings indicate observed correlations (r), correlations corrected for range restriction ($r_c$), and correlations corrected for both range restriction and reliability of the criterion ($r_{fc}$). Statistical significance was tested only for the observed correlations. N = 4,992; $^a p < .05$; $^b p < .01$

## Discussion

This study examined the relations between two cognitive functioning tests and pilot training outcome. Overall, the results were consistent with prior studies of the relations between cognitive ability and pilot training outcomes (e.g., Carretta & Ree, 2003; Hunter & Burke, 1995; Martinussen, 1996; Zierke, 2012). The MAB and MicroCog both assess *g* (Chappelle et al., 2010), as does the AFOQT. The lower validities for the MAB and MicroCog compared to USAF pilot aptitude tests (i.e., AFOQT and PCSM) was likely due to additional factors measured by the pilot aptitude tests. A joint confirmatory factor analysis of the AFOQT and MAB revealed that each test had a hierarchical structure (Carretta, Retzlaff, Callister, & King, 1998). The higher-order factor in the AFOQT has been identified as general cognitive ability (*g*) (Drasgow et al., 2010). The correlation between the higher-order factors from the two tests was .98 indicating that both measured *g*. Although both tests measure *g*, and include verbal, spatial, and perceptual speed content, the AFOQT also includes tests of aviation knowledge not found in the MAB (Carretta et al., 1998). It is likely that the MicroCog does not assess such unique factors either (Chappelle et al., 2010) . The higher validities for the AFOQT Pilot and PCSM composites compared with the MAB and MicroCog is likely due to their measurement of additional factors shown to be related to pilot training performance (e.g., aviation knowledge/experience and psychomotor) that are not included in the MAB and MicroCog, which are primarily designed for clinical assessment. Aviation knowledge and experience may be an indirect measure of motivation.

Nevertheless, the MAB and MicroCog scores demonstrated predictive validity against most of the training criteria. For example, after correction for both range restriction and reliability of the criteria, the MAB FSIQ score was correlated .564 with academic grades, .287 with daily flying grades, .262 with check flight grades, and .363 with class rank. The MicroCog also showed generally significant results when compared against training criteria, but had lower validities than the MAB after correction.

Neither the MAB nor the MicroCog, however, was an effective predictor of graduation versus DOR elimination. This result was expected as DOR elimination is affected by both ability and motivation and neither of the tests assesses motivation. It should be noted that these cognitive functioning tests were not administered with the primary purpose of predicting training outcomes. Rather, their purpose was to baseline cognitive functioning for potential future ideographic comparisons.

As with other occupations (Schmidt & Hunter, 1998), pilot training performance is affected by both ability (can do) and motivation (will do) factors. Cognitive aptitude tests measure the "can do" component of achievement, while factors such as prior aviation experience and specialized job-related knowledge sought by the applicant, and personality, measure the "will do" component. As no USAFSAM cognitive functioning tests directly assess aviation motivation, a future study will examine personality and its incremental validity in the prediction of flying training performance when used in combination with measures of cognitive ability.

## References

Boyd, J. E., Patterson, J. C., & Thompson, B. T. (2005). Psychological test profiles of USAF pilots before training vs. type aircraft flown.

*Aviation, Space, and Environmental Medicine, 76,* 463-468.
Carretta, T. R. (2005). *Development and validation of the Test of Basic Aviation Skills (TBAS),* AFRL-HE-WP-TR-2005-0172. Air Force Research Laboratory, Human Effectiveness Directorate, Wright-Patterson AFB, OH.
Carretta, T. R. (2008). *Predictive validity of the Air Force Officer Qualifying Test for USAF air battle manager training performance*, AFRL-RH-WP-TR-2009-0007. Air Force Research Laboratory, Human Effectiveness Directorate, Wright-Patterson AFB, OH.
Carretta, T. R. (2010). Predictive validity of the Air Force Officer Qualifying Test for non-rated officer specialties. *Military Psychology, 22,* 450-64.
Carretta, T. R. (2011). Pilot Candidate Selection Method: Still an effective predictor of US Air Force pilot training performance. *Aviation Psychology and Applied Human Factors, 1,* 3-8.
Carretta, T. R. (2013). Predictive validity of pilot selection instruments for remotely piloted aircraft training outcome. *Aviation, Space, and Environmental Medicine, 84,* 1-7.
Carretta, T. R., & Ree, M. J. (1995). Air Force Officer Qualifying Test validity for predicting pilot training performance. *Journal of Business and Psychology, 9,* 379-88.
Carretta, T. R., & Ree, M. J. (2003). Pilot selection methods. in P. S. Tsang & M. A. Vidulich, eds., *Human factors in Transportation: Principle s and Practice of Aviation Psychology,* pp. 357-396. Mahwah, NJ: Erlbaum.
Carretta, T. R., Retzlaff, P. D., Callister, J. D., & King, R. E. (1998). A comparison of two US Air Force pilot aptitude tests. *Aviation, Space, and Environmental Medicine, 69,* 931-935.
Chappelle, W., McDonald, K., Thompson, W., McMillan, K., & Marley, M. (2010). *Multiple Aptitude Battery-II normative intelligence test data that distinguish U.S. Air Force AC-130 gunship sensor operators,* AFRL-SA-BR-TR-2010-0006. U.S. Air Force School of Aerospace Medicine, Brooks City-Base, TX.
Chappelle, W., Ree, M. J., Barto, E. L., Teachout, M. S., & Thompson, W. T. (2010). *Joint use of the MAB-II and MicroCog for improvements in the clinical and neuropsychological screening and aeromedical waiver process of rated USAF pilots,* AFRL-SA-BR-TR-2010-0002. US Air Force School of Aerospace Medicine, Brooks City-Base, TX.
Cohen, J. (1983). The cost of dichotomization. *Applied Psychological Measurement, 7,* 249-253.
Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.
Conoley, J. C., & Kramer, J. J. (1989) *The tenth mental measurements yearbook*. Lincoln, NE: The University of Nebraska Press.
Costa, P. T., & McCrae, R. R. (1985). *The NEO Personality Inventory Manual.* Psychological Assessment Resources, Odessa, FL.
Drasgow, F., Nye, C. D., Carretta, T. R., & Ree, M. J. (2010). Factor structure of the Air Force Officer Qualification Test Form S: Analysis and comparison with previous forms. *Military Psychology, 22,* 68–85.
Howse, W. R., & Damos, D. L. (2011). *A bibliographic database for the history of pilot training selection*, AFCAPS-FR-2011-0010. Air Force Personnel Center, Randolph AFB, TX.
Hunter, D. R., & Burke, E. F. (1995). *Handbook of pilot selection.* Brookfield, VT: Avebury Aviation.
Hunter, J. E., & Schmidt, F. L. (2004). *Methods of meta-analysis.* Thousand Oaks, CA: Sage.
Jackson, D. N. (1984). *Multidimensional Aptitude Battery: Manual*. Research Psychologists Press, Port Huron, MI.
Jackson, D. N., (1998). *Multidimensional Aptitude Battery-II: Manual.* SIGMA Assessment Systems, Port Huron, MI.
King, R. E., Barto, E., Ree, M. J., & Teachout, M. S. (2011). *Compilation of pilot personality norms*, AFRL-SA-WP-TR-2011-0008. U.S. Air Force School of Aerospace Medicine, Wright-Patterson AFB, OH.
King, R. E., Barto, E., Ree, M. J., Teachout, M. S., & Retzlaff, P. D. (2011). *Compilation of pilot cognitive ability norms,* AFRL-SA-WP-TR-2012-0001. U.S. Air Force School of Aerospace Medicine, Aeromedical Research Department, Wright-Patterson Air Force Base, OH.
Kuncel, N. R., Hazlett, S. A., & Ones, D. S. (2001). A comprehensive meta-analysis of the predictive validity of the Graduate Record Examinations: Implications for graduate student selection and performance. *Psychological Bulletin, 127,* 162-181.
Kuncel, N. R., Hazlett, S. A., & Ones, D. S. (2004). Academic performance, career potential, creativity, and job performance: Can one construct predict them all? *Journal of Personality and Social Psychology, 86,* 148-161.
Lawley, D. N. (1943). A note on Karl Pearson's selection formulae. *Proceedings of the Royal Society of Edinburgh, 62, (Section A, Part 1),* 28-30.
Lipsey, M. W., & Wilson, D. B. (2000). *Practical Meta-Analysis*. Sage, Thousand Oaks, CA.
Martinussen, M. (1996). Psychological measures as predictors of pilot performance: A meta-analysis. *International Journal of Aviation Psychology, 6,* 1-20.
Paullin, C., Katz, L., Bruskiewicz, K. T., Houston, J., & Damos, D. (2006). *Review of aviator selection*, Technical Report 1183. U.S. Army Research Institute for the Behavioral and Social Sciences, Arlington, VA.
Olea, M. M., & Ree, M. J. (1994). Predicting pilot and navigator criteria: Not much more than *g*. *Journal of Applied Psychology, 79,* 845-51.
Powell, D. H., Kaplan, E. F., Whitla, D., Weintraub, S., Caitlin, R., & Funkenstein, H. H. (1993). *MicroCog: Assessment of cognitive functioning (Version 2.1) Manual*. Psychological Corporation, San Antonio, TX.
Ree, M. J., & Carretta, T. R. (1996). Central role of *g* in military pilot selection. *International Journal of Aviation Psychology, 6,* 111-123.
Roberts, H. E., & Skinner, J. (1996). Gender and racial equity of the Air Force Officer Qualifying Test in officer training school selection decisions. *Military Psychology, 8,* 95-113.
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.
U.S. Air Force (2011). *Medical examinations and standards,* Air Force Instruction 48-123. Washington, DC: Department of the Air Force.
Vanderploeg, R. D., ed. (2000). *Clinician's guide to neuropsychological assessment,* 2nd ed. Mahwah, NJ: Erlbaum.
Wallbrown, F. H., Carmin, C. N., & Barnett, R. W. (1988). Investigating the construct validity of the Multidimensional Aptitude Battery. *Psychological Reports*, *62*, 871-878.
Wallbrown, F. H., 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. (1981). *Wechsler Adult Intelligence Scales – Revised (WAIS-R).* The Psychological Corporation: NY.
Weeks, J. L., & Zelenski, W. E. (1998). *Entry to USAF undergraduate flying training,* AFRL-HE-AZ-TR-1998-0077. Brooks AFB, TX: Air Force Research Laboratory, Training Effectiveness Branch, Warfighter Training Research Division.
Zierke, O. (2012). *Predictive validity in aviation psychology: Really not much more than g?* Paper presented at the 30th Meeting of the European Association of Aviation Psychology. Villasimius, Sardinia, Italy.