Compilation of Cognitive and Personality Norms for Military Aviators

Thomas R. Carretta; Raymond E. King; Malcolm James Ree; Mark S. Teachout; Erica Barto

INTRODUCTION: The assessment of individuals on abilities or other characteristics is based on comparison to a representative sample. General population norms provide an appropriate reference group when the distribution of scores in the sample can be expected to be similar to those for the general population (e.g., comparing high school students at a particular school to national high school norms on a college entrance test). Specialized norms are needed, however, when subsets of the population differ from the population at large. Military pilot trainees represent a special population; they are highly screened on cognitive ability and other characteristics thought to be related to job performance. Other characteristics (e.g., personality) are thought to be "self-selected," resulting in distinctive profiles. Normative tables were developed for U.S. Air Force pilot trainees for two widely used tests, the Multidimensional Aptitude Battery-II (MAB-II) and NEO Personality Inventory-Revised (NEO PI-R).

METHODS: The MAB-II and NEO PI-R were administered to large samples of USAF cadets, ROTC students, and officers selected for pilot training.

RESULTS: The mean MAB-II full-scale IQ was about 1.5 SD above the adult population norm and was much less variable, supporting the need for specialized norms. Tables showing the percentile equivalents are provided for use by clinicians.

DISCUSSION: Use of these tables, in addition to, or in lieu of, commercially published norms, will prove helpful when clinical psychologists perform assessments on pilots; in particular when evaluating them for return-to-duty status following a disqualifying condition that may have affected cognitive functioning or emotional stability.

KEYWORDS: occupational norms, pilot aptitude, psychological assessment, specialized norms.

Carretta TR, King RE, Ree MJ, Teachout MS, Barto E. *Compilation of cognitive and personality norms for military aviators.* Aerosp Med Hum Perform. 2016; 87(9):764–771.

The assessment of human characteristics is based on comparing an individual to a representative sample. Such a procedure allows comparisons to be made between individuals on the distribution of values of a particular characteristic of interest, and is commonplace in organizations (e.g., deciding who to hire or promote). Several texts on psychological testing provide recommendations on the use of test norms.[2,14] Further, professional standards recommend that when norms are used they should refer to clearly defined populations.[1,33] A consistent point in the testing literature is that a relevant normative sample should be used when interpreting the scores for an individual. Certain subsets of the general population vary dramatically from the population at large on several attributes that may be related to occupational performance, including cognitive and physical ability, personality and temperament, and specialized training, job knowledge, and skills. For example, groups can be distinguished on level of academic achievement – high school dropout or graduate, holder of a technical, college, or advanced degree. Moreover, although differences may be found between occupational groups on cognitive ability[20] and personality profiles,[34] individual differences also may be found within occupational groups.

Aviators are a highly selected and distinguished occupational group, with United States Air Force (USAF) pilots at or near the pinnacle. The stakes are high. In this high-risk profession, errors can lead to significant costs in human life, international relations, and national security. Further, human

From the U.S. Air Force School of Aerospace Medicine, Wright-Patterson AFB, OH.

This manuscript was received for review in December 2015. It was accepted for publication in April 2016.

Address correspondence to: Thomas R. Carretta, Ph.D., Air Force Research Laboratory, 711 HPW/RHCI, Wright-Patterson AFB, OH; thomas.carretta@us.af.mil.

DOI: 10.3357/AMHP.4545.2016

error is often a causal factor in pilot training and mission-related aviation mishaps. Due to the high costs of military aviation training and the high-risk nature of military flying, an understanding of pilot cognitive and personality characteristics is a critical part of the selection process.[6] It is also important in the aeromedical waiver process for decisions about return to flying duties after the occurrence of a medically disqualifying injury or illness.[9,22]

The U.S. Air Force collects cognitive and personality test data prior to entry into pilot training.[25] Currently, the battery includes two widely used measures, the Multidimensional Aptitude Battery-2nd Edition (MAB-II)[19] and the NEO Personality Inventory-Revised (NEO PI-R).[13] These premorbid measures are very useful in the comprehensive clinical and neuropsychological evaluations that occur at the USAF School of Aerospace Medicine Aeromedical Consultation Service when pilots are being considered for return to flying duties after a medically disqualifying diagnosis. The archived scores on these measures can be compared to the pilot's current functioning when seeking a waiver to the medical standards.[36]

Specifically, how do baseline psychological data collected aid in future clinical assessments? A baseline of scores for each pilot candidate is created during Medical Flight Screening (MFS), prior to entering training to serve, as an accurate assessment of their premorbid cognitive and personality functioning. These baseline scores are archived to be used for comparison, should the pilot need to undergo a psychological or neuropsychological evaluation during or after training. During the clinical assessment to determine fitness to return to flying duties, an aeromedical psychologist must determine whether there are any changes in a pilot's or pilot candidate's cognitive or personality functioning that would raise concerns if they were to return to flying.

Clinically assessing aviators with a history of a psychiatric illness or a neurological insult can be difficult, as these typically high-functioning people retain abilities well above the general population, even when relatively impaired, especially when occupationally impaired. The risk of returning an impaired pilot to flying duties must be minimized. Because subtle changes in a pilot's cognitive functioning can cause disqualification from flying, aeromedical evaluations are very sensitive. Therefore, having an accurate baseline of the cognitive functioning and personality of pilots is critical to their future, the aeromedical waiver process, and mission readiness. Comparing an individual to their own baseline data is known as an ideographic assessment. In contrast, traditional assessments that compare individuals to collected norms are known as nomothetic assessments. Both types of assessment are useful. In either ideographic or nomothetic assessments, occupationally specialized norms, as opposed to general population, norms are needed.

The information in test manuals usually is based on a representative sample of the general population. Clinical psychological and neuropsychological evaluation of USAF pilots, however, requires occupation-specific normative data. These pilots are not representative of the general population but are a highly screened and selected group. Therefore, compiling pilot norms may be useful to augment those compiled on the general population. A previous study involving 5617 USAF student pilots[35] showed large differences between student pilot and adult population norms, supporting the need for occupation-specific norms.

In several empirical studies, mean score differences in cognitive ability across military[7] and civilian jobs have been observed.[17,20,32] Mean differences in personality across jobs also have been observed.[30] In addition to mean differences, it has been observed that the variability of cognitive ability among individuals within occupations tends to be less than that seen in the general population and varies across occupations.[20] Several studies have discussed the importance of using occupation-specific norms when corrections for range restriction in predictive validation research involving measures of ability[16,29,31] and personality.[30]

The problem with using nationally representative norms with military pilots is that the scores for pilots are usually very high. Scores can fall toward the mean of the nationally representative sample, suggesting that no clinical issue exists in the event of a medically disqualifying injury or illness. However, given the generally extreme scores of USAF pilots, changes in scores toward the nationally normed mean could indicate a serious decrement. This decrement, although reducing the pilot to the average range when compared to the general population, may be an indication that flying duties will not be performed safely and effectively.

For example, say a pilot achieves an MAB-II performance intelligence quotient (PIQ) of 105 after a closed head injury. While this value is in the normal range when compared to the general population (IQs are normed to have a mean of 100 and a standard deviation of 15), it is low (as will be shown) compared to pilots and may be indicative of a loss of neurocognitive functioning. Having MAB-II IQ scores obtained from this individual before injury is invaluable. Similarly, a psychologically well-adjusted pilot with a 95th percentile score on emotional stability that falls to the 60th percentile may be showing signs of a psychopathological condition or evidence of deteriorating interpersonal adjustment. Moreover, the change from high to average may forecast unsuccessful pilot behavior. Occupation-specific norms may lead to improved psychological evaluation and appropriate decisions regarding flying status and other duties.

The purpose of the current study was to update the cognitive norms published by Thompson et al. (35) and to make available aviator-specific personality norms. Use of these tables, in addition to, or in lieu of, commercially published norms, will prove helpful when psychologists perform clinical assessments on pilots. This study reports on two widely-used tests, the MAB-II (cognitive) and NEO PI-R (personality).

It should be noted that this type of undertaking is not new. Fine and Hartman[15] compiled norms on the Minnesota Multiphasic Personality Inventory (MMPI)[4] for USAF pilots. These norms subsequently were transformed into a profile sheet and used during aeromedical assessments at the Aeromedical Consultation Service (ACS) until the ACS adopted

Table I. Means and SDs for MAB-II IQ Scores and Subtests.

	MEN		WOMEN		ALL	
	N = 22,797		*N* = 2192		*N* = 24,989	
SCORE	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**
FSIQ	120.75	6.16	118.85	6.49	120.58	6.19
VIQ	119.35	6.71	118.07	6.52	119.24	6.69
PIQ	119.43	8.32	117.12	8.18	119.23	8.31
Information	66.69	6.25	65.02	6.51	66.54	6.27
Comprehension	59.53	4.27	59.26	4.19	59.50	4.26
Arithmetic	61.29	6.74	58.82	6.25	61.07	6.70
Similarities	60.09	5.28	60.55	4.66	60.10	5.22
Vocabulary	59.60	7.18	59.84	7.43	59.62	7.20
Digit Symbol	65.75	6.77	67.41	5.93	65.89	6.69
Picture Completion	59.96	6.35	56.69	6.30	59.67	6.34
Spatial	60.20	7.00	57.71	7.30	59.98	7.03
Picture Arrangement	52.05	7.38	51.13	7.24	51.97	7.37
Object Assembly	60.86	5.70	60.01	6.61	60.78	5.78

the revised version of the MMPI (MMPI-2)[5] in 1991. As noted above, Thompson et al.[35] published cognitive norms for USAF pilot trainees.

Not all of the pilot trainees in the ACS database successfully complete training.[35] While individual pilot's baseline data will remain useful for further assessments, aggregate data that includes both training graduates and eliminees may need to be used more cautiously as it may not completely reflect the performance of rated USAF pilots. Those interested in differences in test scores for successful and unsuccessful pilot trainees are referred elsewhere.[8,26]

In considering the issue of fitness for duty of aviators, Kay asserted that "there is no excuse for inexactness in the evaluation of individuals in these safety-critical occupations".[21(p.228)] The current analyses attempt to mitigate one obstacle confronted by the professionals who attempt to assess those who are typically high functioning.

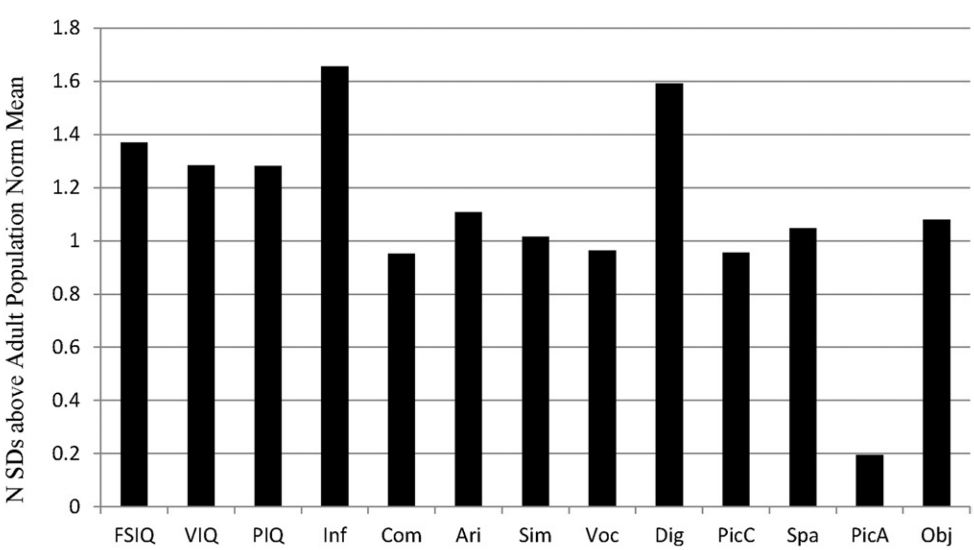

Fig. 1. Number of standard deviations pilot trainees differed from adult population norms for the MAB-II composites and subtests. Positive values indicate the mean for pilots is above the adult population mean. Abbreviations are: Full-Scale IQ (FSIQ), Verbal IQ (VIQ), Performance IQ (PIQ), Information (Inf), Comprehension (Com), Arithmetic (Ari), Similarities (Sim), Vocabulary (Voc), Digit Span (Dig), Picture Completion (PicC), Spatial (Spa), Picture Arrangement (PicA), and Object Assembly (Obj).

METHODS

Subjects

The MAB-II and NEO PI-R were administered to large samples of USAF cadets and officers selected for pilot training as part of the Medical Flight Screening (MFS) assessment prior to beginning Specialized Undergraduate Pilot Training. The sample sizes for the MAB-II and NEO PI-R overlap but are not identical. A sample of 25,514 trainees completed the MAB-II while 12,702 completed the NEO PI-R. The samples for the MAB-II and NEO PI-R are different for several reasons. While participation in the cognitive portion of MFS was mandatory, participation in personality testing was voluntary. Also, the tests used in personality testing varied over the course of the MFS program, while the MAB-II was a mainstay in cognitive testing. The two samples were similar in their demographic makeup. All were college graduates or were near completion of college when they were tested. Many had private pilot's licenses or had completed part of training for a private pilot's license including flight hours in a light aircraft. Of those reporting demographic information, most were male (MAB-II: 91.2%: NEO PI-R: 92.9%). All participants were under the age of 40 with a mean age of about 23 yr (MAB-II: M = 22.8 yrs., SD = 2.7 yrs.; NEO PI-R: M = 23.0 yrs., SD = 2.6 yr). Ethnic and racial distributions indicated that most identified themselves as White (MAB-II: 84.2%; NEO PI-R: 88.9%) and that the samples had similar proportions of Hispanics (MAB-II: 4.0%; NEO PI-R: 3.6%) and African-Americans (MAB-II: 2.4%; NEO PI-R: 2.0%). All participants were tested at either the USAF School of Aerospace Medicine at Brooks City-Base, TX, or at the USAF Academy in Colorado Springs, CO. The voluntary informed consent of the subjects used in this study was obtained as required by U.S. Air Force regulations.

Measures

Multidimensional Aptitude Battery. The MAB[18] and MAB-II[19] are broad-based tests of intellectual ability modeled after the Wechsler Adult Intelligence Scale – Revised (WAIS-R;[39]). The 10 MAB and MAB-II subtests are combined to create three summary scores: verbal IQ (VIQ), performance IQ (PIQ), and full-scale IQ (FSIQ). Detailed descriptions of the subtests are available elsewhere.[18,19] The full-scale IQ scores for the MAB and WAIS-R are strongly correlated ($r = 0.91$;[12,18]). The MAB and MAB-II require less than 1.5 h to administer and can be individually or group administered. Each of the subtests has a mean of 50 and standard deviation (SD) of 10. The IQ scores have a mean of 100 and a SD of 15 in the general population. Norms are based on a sampling of nine age groups that were diverse in terms of gender, ethnicity, race, and North American (Canada and United States) geographic location. Test-retest reliability for the IQ scores ranges from 0.94 to 0.98[19] for a retest interval averaging 45 d. Previous research has shown that the FSIQ score measures general cognitive ability in several age groups.[27,28,37,38]

NEO PI–R. The NEO PI-R[13] was designed to measure the Big Five personality domains and the facets or traits that underlie each domain. The five domains are Neuroticism (N), Extraversion (E), Openness to Experience (O), Agreeableness (A), and Conscientiousness (C). Each domain consists of six subscales called facets. These domains and facets provide a comprehensive measurement of adult personality.

The NEO PI–R was developed with the goal of being a multipurpose personality inventory useful for predicting many criteria such as behaviors related to illness, career interests, psychological health, and coping styles.[13] Examinees are required to respond to 240 statements using a 5-point Likert-type scale, ranging from 1 (strongly disagree) to 5 (strongly agree). Descriptions of the five domain scales are provided in the test manual.[13] Internal consistency reliabilities (coefficient alpha) in a sample of 1539 men and women in a large organization ranged from 0.86 (Agreeableness) to 0.92 (Neuroticism) for the domain scores and from 0.56 to 0.81 for the 30 facets.[13] For the current study, the normative sample for adults served as the normative reference and the test was administered and scored via computer.[13]

Analyses

Means and SDs were computed for men, women, and the combined sample for the MAB-II IQ scores and subtests and for the NEO PI-R domains and facets. The purpose of including combined sample analyses, despite the predominance of men (about 92%), was to assist clinicians in the United States who must comply with the requirements of the Civil Rights Act of 1991.[10] Norms based on gender, race, religion, ethnicity, or national origin are prohibited by Section 106 of that Federal law for use in employment selection. Percentile tables were created for men,

Table II. Percentile Equivalents for MAB-II IQ Scores.

	FSIQ			VIQ			PIQ		
Raw Score	**Men**	**Women**	**All**	**Men**	**Women**	**All**	**Men**	**Women**	**All**
<93	1	1	1	1	1	1	1	1	1
93	1	1	1	1	1	1	1	1	1
94	1	1	1	1	1	1	1	1	1
95	1	1	1	1	1	1	1	1	1
96	1	1	1	1	1	1	1	1	1
97	1	1	1	1	1	1	1	2	1
98	1	1	1	1	1	1	2	2	2
99	1	1	1	1	1	1	2	3	2
100	1	1	1	1	1	1	2	3	2
101	1	1	1	1	1	1	3	4	3
102	1	1	1	1	1	1	3	5	4
103	1	2	1	2	2	2	4	6	4
104	2	2	2	2	2	2	5	7	5
105	2	3	2	3	3	3	6	9	6
106	3	3	3	3	4	4	7	10	7
107	3	4	3	5	5	5	8	13	9
108	4	6	4	6	7	6	10	15	11
109	5	8	6	7	9	8	12	17	12
110	7	10	7	9	11	9	14	21	14
111	9	12	9	12	14	12	17	24	17
112	11	16	11	15	18	15	19	26	19
113	13	19	14	18	22	18	22	30	23
114	17	24	17	22	29	23	26	34	27
115	20	28	21	27	34	27	29	38	29
116	24	34	25	33	40	33	33	44	34
117	29	40	30	38	47	38	38	48	39
118	34	46	35	44	54	44	41	52	42
119	39	53	41	49	60	50	46	58	47
120	46	59	47	55	65	56	52	64	53
121	52	65	53	61	70	62	56	68	57
122	58	70	59	67	76	68	61	73	63
123	64	77	65	73	81	73	67	78	68
124	70	81	71	78	84	79	72	83	73
125	76	86	77	83	88	83	76	86	77
126	81	89	82	87	91	87	81	89	81
127	85	93	86	90	93	90	85	92	85
128	89	95	90	93	95	93	87	94	88
129	92	96	93	95	96	95	91	96	91
130	95	97	95	96	97	96	93	97	93
131	97	98	97	97	98	97	95	98	95
132	98	99	98	98	99	98	96	98	97
133	99	99	99	99	99	99	98	99	98
134	99	99	99	99	99	99	98	99	98
135	99	99	99	99	99	99	99	99	99
>135	99	99	99	99	99	99	99	99	99

women, and the combined sample to show the percentile corresponding to a particular scaled score for the MAB-II IQ scores and NEO PI-R domain scores. Percentile tables for the MAB-II subtests and NEO PI-R facet scores are not provided due to space limitations, but are publically available elsewhere.[23,24] It should be noted that due to the much higher proportion of men in the samples (MAB-II: 91.2%: NEO PI-R: 92.9%), the combined sample statistics and percentile tables largely reflect the male norms. The combined norms are provided for completeness, but the gender-specific norms may be more informative for researchers and practitioners.

RESULTS

Multidimensional Aptitude Battery

Table I summarizes the means and SDs for the MAB-II IQ and subtest scores for males, females, and the combined sample. Test scores for the pilot trainees were severely range restricted compared to the normative values where the means and SDs are 100 and 15 for the IQ scores. The mean IQs for males, females, and the combined sample were about 120 (about 1.33 SD 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 combined sample of trainees was about 17% of the normative value. For the subtests where the normative population mean and SD are 50 and 10, on average the means for the combined pilot trainee sample were about 1 SD above the population norm, and the variances relative to population values ranged from 18.1% (Comprehension) to 54.4% (Picture Arrangement). The largest differences between the pilot and population norms were for the Information ($M = 66.54$) and Digit Symbol ($M = 65.59$) subtests. The smallest was for Picture Arrangement ($M = 51.97$). The pilot trainee score distributions were clearly different from the general population norms for both mean performance and variability of performance.

Cohen[11] characterizes standardized mean score differences (d) between groups of 0.2 as small, 0.5 as medium, and 0.8 or larger as large. With the exception of Picture Arrangement, all mean score differences between the pilot trainees and the normal population were in the large range. **Fig. 1** graphically illustrates the number of SDs the pilot trainee means differed from those of the adult population.

There were small gender differences in mean MAB-II test scores. Although there are instances where the mean for women was higher than that for men on the subtests, men scored 1-2 points higher on all three IQ scores. The difference of 1.9 points for men and women on the FSIQ score would be only 0.12 SD in the normal population where the SD is 15, but was 0.31 SD in the pilot sample where the SD in the combined sample was only 6.19.

Table II shows the percentile corresponding to a particular scaled score on the MAB-II IQ scores for the male, female, and combined samples. Percentile tables for the MAB-II subtests are available online in King et al.[24] By way of example, a male pilot with a scaled VIQ score of 105 would be in only the 3rd percentile of pilot trainees (1.61 SD below the pilot mean of 120 using the pilot SD of 6.19), but would be above average (63rd percentile) for the “normal” population (0.33 SD above the normal population mean using its mean of 100 and SD of 15).

Table III. Means and SDs for the NEO PI-R Domain and Facet Scores.

	MEN		WOMEN		COMBINED	
DOMAIN/FACET	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**
Neuroticism (N)	46.64	9.30	45.63	9.89	46.57	9,35
Anxiety	47.25	9.32	46.65	9.33	47.21	9.32
Angry/Hostility	48.44	9.86	47.81	9,68	48.40	9.85
Depression	46.49	8.18	45.69	8.78	46.44	8.22
Self-Consciousness	46.98	9.77	46.00	10.24	46.91	9.80
Impulsiveness	48.15	10.61	48.16	10.57	48.16	10.61
Vulnerability	42.70	8.62	41.99	8.42	42.65	8.60
Extraversion (E)	57.47	9.65	56.97	9.97	57.41	9.68
Warmth	52.09	9.64	52.09	10.28	52.08	9.68
Gregariousness	55.66	10.13	54.97	11.19	55.59	10.22
Assertiveness	58.25	9,36	59.25	9.17	58.31	9.34
Activity	57.81	8.72	58.79	8.35	57.86	8.70
Excitement-Seeking	62.02	8.27	62.21	8.37	62.01	8.29
Positive Emotions	54.38	10.06	56.79	9.70	54.53	10.05
Openness to Experience (O)	50.20	10.09	55.63	9.82	50.59	10.16
Fantasy	52.09	10.52	55.21	10.15	52.32	10.52
Aesthetics	48.81	10.46	52.02	10.75	49.05	10.51
Feelings	52.29	11.02	54.06	9.67	52.40	10.94
Actions	51.91	10.51	55.20	10.44	52.15	10.54
Ideas	53.94	10.39	57.74	9.71	54.22	10.38
Values	46.48	10.17	52.62	10.55	46.91	10.32
Agreeableness (A)	44.18	10.57	43.14	10.41	44.12	10.56
Trust	49.50	10.47	50.26	11.41	49.57	10.55
Straightforwardness	47.91	10.32	47.44	9.96	47.88	10.31
Altruism	52.72	9.80	51.62	10.67	52.63	9.86
Compliance	45.51	11.28	44.10	10.55	45.41	11.23
Modesty	47.42	10.62	46.83	11.58	47.38	10.69
Tender-Mindedness	46.64	9.93	44.92	10.74	46.50	9.99
Conscientiousness (C)	54.93	10.15	54.34	10.65	54.88	10.19
Competence	55.85	9.17	55.76	9.53	55.84	9.20
Order	50.67	10.31	50.68	10.86	50.68	10.34
Dutifulness	52.97	9.09	51.38	9.86	52.86	9.16
Achievement	59.22	9.21	59.56	9.75	59.22	9.25
Self-Discipline	52.64	9.56	52.51	9.77	52.63	9.57
Deliberation	50.67	10.27	50.74	9.68	50.68	10.22

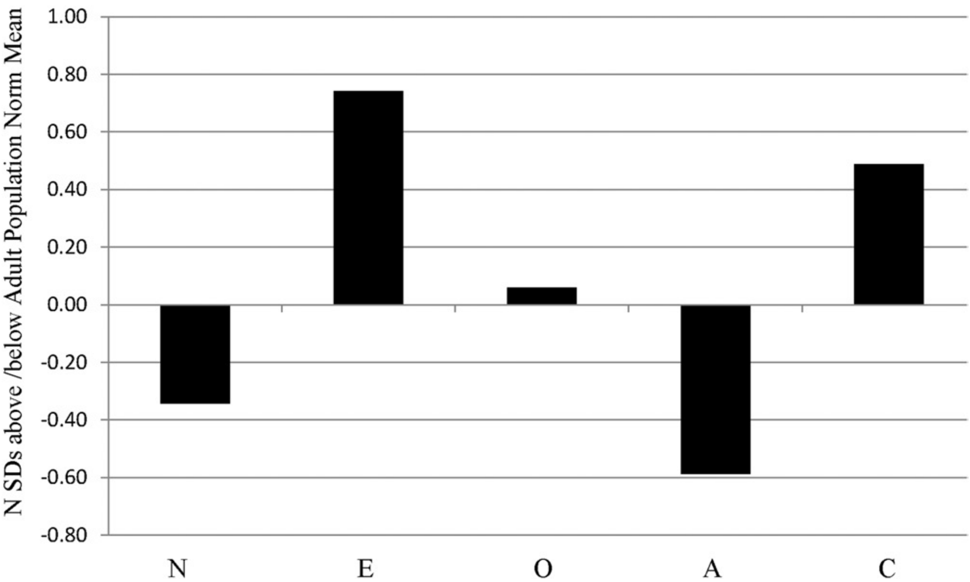

Fig. 2. Number of standard deviations pilot trainees differed from adult population norms for the NEO PI-R domains. Positive (negative) values indicate the mean for pilots was above (below) the adult population mean. Abbreviations are: Neuroticism (N), Extraversion (E), Openness (O), Agreeableness (A), and Conscientiousness (C).

NEO PI-R

Table III summarizes the means and SDs for the NEO PI-R domain and facet scores. Pilot trainees differed from population norms by about 0.5 SD on 4 of the 5 domain scores. Compared to the population norms (M = 50, SD = 10), the USAF pilot trainee means were elevated for Extraversion (E) (M = 57.41) and Conscientiousness (C) (M = 54.88), but lower for Neuroticism (N) (M = 46.57) and Agreeableness (A) (M = 44.12). They were about the same for Openness (O) (M = 50.59). **Fig. 2** displays the differences between the mean scores for the pilot trainees versus the general population norms for the domain scores. On the facet level, the largest deviations from the population norms occurred for N - Vulnerability (M = 42.65), E – Assertiveness (M = 58.31), Activity (M = 57.86), and Excitement Seeking (M = 62.01), and C – Achievement (M = 59.22). It is interesting to note that whereas pilot cognitive scores were much less variable than the adult population norms, pilot personality scores [SD for domains ranged from 9.35 (N) to 10.56 (A)] showed similar variability to those in the adult population where the SD is 10.

With the exception of Openness to Experience, where the mean was moderately higher for women ($d = 0.54$), all domain score gender differences were small.[11] Due to their length, tables showing the percentile corresponding to a particular scaled score on the NEO PI-R domains and facets for the male, female, and combined samples could not be provided here. They are available online in King et al.[23]

DISCUSSION

Pilot trainees have elevated cognitive ability scores coupled with lower variability relative to general population adult norms. The severe range restriction on cognitive ability provides support for the need for specialized norms for pilot trainees. This result is not surprising as pilot trainees are directly selected on indicators of cognitive ability such as college entrance exams, completion of a college degree, and scores on cognitive ability tests used for officer commissioning and pilot training qualification. The amount of range restriction for pilot personality norms relative to the adult population is relatively small compared to cognitive ability. The reason personality scores are less range-restricted is that there are no direct pilot selection procedures for personality. Personality may exert influence indirectly through self-selection, completion of college, and commander's ratings. This indirect measurement yields small mean difference and small variability differences compared to the general population.

The elevated mean (120 vs. 100) and lower variability (6.19 vs. 15) for IQ for USAF pilot trainees relative to the adult population has clinical consequences. The population-specific norms (means and SDs) for pilot trainees provide context for the interpretation of test scores and any deviations from the mean IQ. While a pilot with a measured IQ of 108 is approximately 0.5 SD above the general population (using the general population mean and SD as a reference), they are approximately two SDs below the mean pilot IQ (using the pilot mean and SD as a reference).

As previously noted, not all pilot candidates selected for USAF pilot training successfully complete training.[35] About 18% fail for a variety of reasons, including academic, poor flying performance, lack of motivation (self-elimination), and medical. While data for individual pilots will remain useful for future assessments, aggregate data reported in this study, which includes scores for both training graduates and eliminees, may need to be used more cautiously as it may not be completely representative of pilots who successfully completed training. Comparison between pilot training graduates and eliminees show a mean difference in full-scale IQ of less than 2 points. Those who fail USAF pilot training are relatively high on cognitive functioning comparted with the general adult population. Interested readers are referred to publications that address differences in MFS test results between those who were successful in training and those who were not.[8,26] For example, the traits of conscientiousness and confidence are associated with higher class rank while the traits of negativity, anxiety, depression, and affective lability are associated with lower class rank.[26]

There are two approaches to conduct clinical assessments of pilots seeking a waiver for return-to-duty following an injury or illness that may have resulted in impairment. One method is to develop an individual registry for each pilot against which future testing might be compared. The U.S. Air Force uses this approach, collecting baseline cognitive and personality data during Medical Flight Screening. Pilots are retested during a return-to-flying status evaluation and their performance is compared with baseline scores collected prior to pilot training to determine whether any changes have occurred. Individualized (pre/post) comparisons result in more reliable return-to-flying duties decisions as pilots typically evidence very high cognitive functioning, especially compared to general population norms, and may remain so even after a neurological event or injury.[24]

Although comparison to an individual's premorbid baseline scores is ideal, such an intraindividual comparison is not always possible. Baseline data may not have been captured, or access to the baseline data may not be immediately available. In these situations it is far preferable to use occupationally specific norms as opposed to those complied from the general population. For example, clinicians should use occupational (aviator) norms, although cautiously, when assessing aviators or aviator applicants from outside the USAF for whom baseline norms are not available because they were not captured. Similarly, it is more clinically advisable and justifiable to compare the now relatively rare USAF aviator for whom baseline data are not available, for whatever reason, to their USAF aviator peers rather than to the general population. As noted by Barto, Chappelle, King, Ree, and Teachout,[39(p.12)] "Air Force pilots are a highly selected group whose scores are quite different than the general population, suggesting that clinical evaluations might be quite different if only the normative population was used as a comparison group."

Clinical methods for the neuropsychological assessment of pilots without premorbid cognitive and personality assessments focus on the comparison of an individual pilot's test scores to norms for a reference group. As has been demonstrated, specialized reference group norms are needed for pilots in order to make accurate assessments as they differ markedly from the adult population.

ACKNOWLEDGMENTS

The opinions expressed are those of the authors and not necessarily those of the U.S. Government, U.S. Department of Defense, Federal Aviation Administration, or the U.S. Air Force.

The establishment of the Neuropsychiatric Enhanced Flight Screening (N-EFS), later titled Medical Flight Screening (MFS) program, owes its existence to the early (1993) support of Dr. Penny Giovanetti and Dr. Gerald Saboe. Dr. Raymond E. King was assigned to the Aerospace Medicine Consultation Service, Brooks AFB, TX, at the commencement of N-EFS and served as its first principal investigation and was subsequently a principal investigator on several MFS research projects. The authors extend gratitude to all the principal investigators and support personnel who have kept N-EFS/MFS viable for well over 20 years. Above all, we thank the thousands of USAF student pilots who have participated.

Authors and affiliations: Thomas R. Carretta, Ph.D., Air Force Research Laboratory, Wright-Patterson AFB, OH; Rayomd E. King, Psy.D., Federal Aviation Administration, Washington, DC; Malcolm James Ree, Ph.D., Our Lady of the Lake University, San Antonio, TX; Mark S. Teachout, Ph.D., University of the Incarnate Word, San Antonio, TX; and Erica Barto, M.S., Operational Technologies, Inc. San Antonio, TX.

REFERENCES

1. American Psychological Association. Standards for educational and psychological testing. Washington (DC): American Psychological Association; 1999.
2. Anastasi A, Urbana S. Psychological testing. Upper Saddle River (NJ): Prentice-Hall; 1997.
3. Barto E, Chappelle W, King RE, Ree MJ, Teachout MS. The NEO PI-R as a premorbid baseline measure. Brooks City Base (TX): U.S. Air Force School of Aerospace Medicine; 2011; AFRL-SA-WP-TR-2011-0001.
4. Buchanan RD. The development of the Minnesota Multiphasic Personality Inventory. J Hist Behav Sci. 1994; 30:148–161.
5. Butcher JN, Dahlstrom WG, Graham JR, Tellegen A, Kaemmer B. The Minnesota Multiphasic Personality Inventory-2 (MMPI-2): manual for administration and scoring. Minneapolis (MN): University of Minnesota Press; 1989.
6. Carretta TR, Ree MJ. Pilot selection methods. In: Kantowitz BH series editor; Tsang PS, Vidulich MA, volume editors. Human factors in transportation; principles and practice of aviation psychology. Mahwah (NJ): Erlbaum; 2003:357–396.
7. Carretta TR, Ree MJ, Teachout MS. Effects of training on variability of performance within and across jobs. Int J Sel Assess. 2016; 24:71–76.
8. Carretta TR, Teachout MS, Ree MJ, Barto EL, King RE, Michaels CF. Consistency of the relations of cognitive ability and personality traits to pilot training performance. Int J Aviat Psychol. 2014; 24:247–264.
9. Chappelle W, Ree MJ, Barto EL, Teachout MS, Thompson WT. Joint use of the MAB-II and MicroCog for improvements in the clinical and neuropsychological screening and aeromedical waiver process of rated USAF pilots. Brooks City Base (TX): U.S. Air Force School of Aerospace Medicine; 2010; AFRL-SA-BR-TR-2010-0002.
10. Civil Rights Act of 1991; § 109, 42 U.S.C.; § 2000e *et seq*; 1991.
11. Cohen J. Statistical power analysis for the behavioral sciences, 2nd ed. Mahwah (NJ): Erlbaum; 1988.
12. Conoley JC, Kramer JJ. The tenth mental measurements yearbook. Lincoln (NE): The University of Nebraska Press; 1989.
13. Costa PT, McCrae RR. Professional Manual Revised NEO Personality Inventory (NEO PI-R) and NEO Five Factor Inventory (NEO-FFI). Odessa (FL): Psychological Assessment Resources; 1992.
14. Crocker L, Algina J. Norms and Standard Scores. In: Introduction to modern and classical test theory, Chapter 19. New York (NY): Hartcourt, Brace, & Jovanovich; 1986.
15. Fine PM, Hartman BO. Psychiatric strengths and weaknesses of typical Air Force pilots. Brooks AFB (TX): U.S. Air Force School of Aerospace Medicine, Aerospace Medical Division; 1968, SAM-TR-68-121.
16. Hoffman CC. Applying range restriction corrections using published norms: three case studies. Pers Psychol. 1995; 48:913–923.
17. Hunter JE. Test validation for 12,000 jobs: An application of job classification and validity generalization to the General Aptitude Test Battery. Washington (DC): U.S. Department of Labor; 1983; U.S. Employment Service Test Research Report 43.
18. Jackson DN. Multidimensional Aptitude Battery: manual. Port Huron (MI): Research Psychologists Press; 1984.
19. Jackson DN. Multidimensional Aptitude Battery-II: manual. Port Huron (MI): SIGMA Assessment Systems; 1998.
20. Jensen AR. The g factor: the science of mental ability. Westport (CT): Praeger-Greenwood; 1998.
21. Kay GG. Guidelines for the psychological evaluation of air crew personnel. Occupation Med. 2002; 17:227–245.

22. King RE. Psychological testing for mental health screening, suitability determinations, and archival purposes to improve safety and reduce costs. In: Droog A, Heese M editors. Proceedings of the 30th Conference of the European Association for Aviation Psychology, Sardinia, Italy, September 2–28, 2012 . Brussels (Belgium): EAAP; 2012: 51–56.
23. King RE, Barto E, Ree MJ, Teachout MS. Compilation of pilot personality norms. Wright-Patterson AFB (OH): School of Aerospace Medicine, Aeromedical Research Division; 2011; AFRL-SA-WP-TR-2011-0008.
24. King RE, Barto E, Ree MJ, Teachout MS, Retzlaff P. Compilation of pilot cognitive ability norms. Wright-Patterson AFB (OH): USAF School of Aerospace Medicine, Aeromedical Research Division; 2011; AFRL-SA-WP-TR-2012-0001. Available at http://www.dtic.mil/dtic/tr/fulltext/u2/a554594.pdf
25. King RE, Flynn CF. Defining and measuring the "Right Stuff": neuropsychiatrically enhanced flight screening (N-EFS). Aviat Space Environ Med. 1995; 66(10):951–956.
26. King RE, Retzlaff P, Barto E, Ree MJ, Teachout MS. Pilot personality and training outcomes. Wright-Patterson AFB (OH): USAF School of Aerospace Medicine, Aeromedical Research Division; 2012; AFRL-SA-WP-TR-2012-0013.
27. Kranzler JH. The construct validity of the Multidimensional Aptitude Battery: a word of caution. J Clin Psychol. 1991; 47(5):691–697.
28. Lee MS, Wallbrown FH, Blaha J. Note on the construct validity of the Multidimensional Aptitude Battery. Psychol Rep. 1990; 67(3):1219–1222.
29. Linn RL, Harnisch DL, Dunbar SB. Corrections for range restriction: an empirical investigation of conditions resulting in conservative corrections. J Appl Psychol. 1981; 66(6):655–663.
30. Ones DS, Viswesvaran C. Job-specific applicant pools and national norms for personality scales: Implications for range-restrictions corrections in validation research. J Appl Psychol. 2003; 88(3):570–577.
31. Sackett PR, Ostgard DJ. Job-specific applicant pools and national norms for cognitive ability tests: implications for range-restriction corrections in validation research. J Appl Psychol. 1994; 79(5):680–684.
32. Schmidt FL, Ones DS, Hunter JE. Personnel selection. In: Rosenzweig MR, Porter LW, editors. Annual review of psychology. Palo Alto (CA): Consulting Psychologists Press; 1992: 627–670.
33. Society for Industrial and Organizational Psychology. Principles for the validation and use of personnel selection procedures, 4th ed. Bowling Green (OH): SIOP; 2003.
34. Tett RP, Fitzke JR, Wadlington PL, Davies SA, Anderson MG, Foster J. The use of personality test norms in work settings: effects of sample size and relevance. J Occup Organ Psychol. 2009; 82(3):639–659.
35. Thompson WT, Orme DR, Zazeckis TM. Neuropsychological evaluation of aviators: need for aviator-specific norms? Brooks City Base (TX): U.S. Air Force School of Aerospace Medicine; 2004; SAM-FE-BR-TR-2004-0001.
36. United States Air Force Medical examinations and standards, Air Force Instruction 48-123. Washington (DC): Department of the Air Force; 2011.
37. Wallbrown FH, Carmin CN, Barnett RW. Investigating the construct validity of the Multidimensional Aptitude Battery. Psych Rep 1988; 62(3):871–878.
38. Wallbrown FH, Carmin CN, Barnett RW. A further note on the construct validity of the Multidimensional Aptitude Battery. J Clin Psychol. 1989; 45(3):429–433.
39. Wechsler D. Wechsler Adult Intelligence Scales – Revised (WAIS-R). New York (NY): The Psychological Corporation; 1981.