

University of Nebraska - Lincoln

DigitalCommons@University of Nebraska - Lincoln

U.S. Air Force Research

U.S. Department of Defense

2016

Aptitude and Trait Predictors of Manned and Unmanned Aircraft Pilot Job Performance

Laura Barron

Thomas R. Carretta

Mark R. Rose

Follow this and additional works at: <https://digitalcommons.unl.edu/usafresearch>

This Article is brought to you for free and open access by the U.S. Department of Defense at DigitalCommons@University of Nebraska - Lincoln. It has been accepted for inclusion in U.S. Air Force Research by an authorized administrator of DigitalCommons@University of Nebraska - Lincoln.

Aptitude and Trait Predictors of Manned and Unmanned Aircraft Pilot Job Performance

Laura G. Barron
Air Force Personnel Center
Randolph AFB, TX

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

Mark R. Rose
Air Force Personnel Center
Randolph AFB, TX

Previous research has shown that the same aptitude and trait measures that predict success in US Air Force (USAF) manned aircraft pilot training predict remotely-piloted aircraft (RPA) pilot training outcomes with generally similar levels of validity (Carretta, 2013; Rose, Barron, Carretta, Arnold, & Howse, 2014). However, because USAF RPA pilots initially train in manned aircraft, validation of aptitude and traits predicative of RPA pilot success has thus far been limited to RPA pilot training outcomes that actually require manned flight. Hence, thus far there has been no basis for determining the aptitudes and traits predictive of success in environments in which pilots actually fly RPAs. To address this gap, the present study evaluated pre-accession trait (Big Five personality domains) and aptitude (spatial, quantitative, and aviation knowledge) measures as predictors of manned and unmanned aircraft pilot performance on-the-job, as measured by supervisor and senior rater stratification on three years of Officer Performance Reports (OPRs). Results were generally consistent in showing that the same aptitudes, knowledge, and personality traits that predict successful job performance for manned aircraft pilots also predict successful job performance for RPA pilots. However, results also showed pre-accession aviation knowledge to be a stronger predictor of job performance for RPA pilots than for manned aircraft pilots. These findings and their implications for attracting and selecting RPA pilots are discussed.

Keywords: aviator aptitude and personality traits, job performance, remotely piloted aircraft, unmanned aircraft systems

Military and civilian use of remotely piloted aircraft (RPA) (alternately termed unmanned aerial systems [UAS] or, pejoratively, “drones”) has increased exponentially in recent years (Dillingham, 2012; Gertler, 2012). Despite continuing improvements in technology and automation, RPAs do not fly themselves, and there is increasing demand (both in military and civilian applications) for skilled RPA pilots. For example, the US Department of Defense plans a 50 percent increase in RPA combat air patrols from 2015 to 2019 (Everstine, 2015), and it is estimated that more than 100,000 new RPA jobs

in a variety of industries ranging from agriculture to public safety will be created as a result of integrating RPAs into the US national airspace system (Jenkins & Vasigh, 2013).

In the popular press, there has been an assumption that so-called “drone” (RPA/UAS) piloting is fundamentally different from manned aircraft piloting, potentially requiring different types of aptitudes or personal traits. For example, RPA pilots often have been portrayed as “glorified video game players” (see Roughton, 2015) who “sit in lounge chairs at a safe physical distance, thousands of miles from harm’s way” (see Greene Sands & Dove-Steinkamp, 2015, p. 1). A related popular press notion is that “soldiers engaged in such ‘virtual warfare’ are less situationally aware, and also less restrained because of emotional detachment” (see Aerospace Industries Association, 2013, p. 7).

The views expressed are those of the authors and are not necessarily those of the US Government, Department of Defense, or the US Air Force. Correspondence should be addressed to Laura Barron at HQ AFPC/DSYX Strategic Research & Assessment, 550 C St W, Randolph AFB, TX 78150-4747. Contact: Laura.Barron@us.af.mil

In contrast to the popular press, the U.S. Air Force appears not to have presumed that the knowledge, aptitudes, and traits that are important for RPA piloting are substantively different from those relevant for traditional manned aircraft pilots. In fact, it may be more challenging to navigate the RPA because of the absence of some sensations and resulting intuition gained from flying a manned aircraft (e.g., Tvaryanas, 2006), a view commonly shared by RPA pilots with manned aircraft experience. Although the U.S. Air Force has had a separate, dedicated career field for RPA pilots since 2009, the Air Force has explicitly designed RPA pilot training to parallel manned aircraft training. Specifically, the training pipeline requires an introductory flight screening (IFS) phase in manned aircraft (39.3 hours of hands-on flight for RPA candidates) prior to beginning RPA-specific instruction. Hence, ironically, almost all of the RPA pilot training attrition has been based on failures while piloting manned aircraft. Even follow-on training (RPA Instrument Qualification) is based on simulators of traditional manned aircraft.

To the extent that U.S. Air Force RPA pilot training actually requires flight of manned aircraft, one would clearly expect that the same long-standing measures used to predict who will be successful in manned aircraft pilot training would also predict success in RPA pilot training. Indeed, studies have confirmed that the same pilot selection measures (knowledge, aptitude, and trait measures) that predict success in US Air Force (USAF) manned aircraft pilot training show similar levels of predictive validity for RPA pilot training (Carretta, 2013; Rose, Barron, Carretta, Arnold, & Howse, 2014).

Hence the question becomes: are these relationships showing similar knowledge, aptitude, and trait predictors of RPA and manned aircraft pilot training (Carretta, 2013; Rose, Barron, Carretta, Arnold, & Howse, 2014) an artifact of the decision by the USAF to explicitly design the RPA training curriculum to parallel that for manned aircraft pilots? Or, alternately, as the popular press has implicitly suggested, are there actually fundamental differences such that RPA pilot job performance requires a different set of aptitudes or personality traits from those relevant for manned aircraft piloting? For example, is emotional stability less predictive of the performance of RPA pilots than manned aircraft pilots because RPA pilots do not face the stressor of being physically in harm's way? Or, as an alternate example, are spatial aptitudes actually more predictive of the performance of RPA pilots than traditional pilots because RPA pilots are deprived of many of the sensory cues available in manned aircraft? By using a measure of job performance, rather than training

performance, as a criterion for validation, the purpose of this present study is to provide a more appropriate basis for comparison of those aptitudes and traits predictive of success for RPA versus manned aircraft pilots.

A History of Pilot Selection Research based on Predictors of Training Outcomes

Myriad studies have documented the aptitude, knowledge, and (to a lesser extent) trait measures that predict success in pilot training (Campbell, Castaneda, & Pulos, 2010; Hunter & Burke, 1994; Martinussen, 1996). Beginning with the growth of aviation during World War II, pilot training attrition has been high, and the military invested heavily in research on individual differences that could distinguish those recruits that could be readily trained to become pilots (Damos, 2011). Such research has consistently shown that the types of measures the Air Force currently uses in pilot selection—including tests of quantitative and spatial abilities, as well as more aviation-specific knowledge tests—show useful levels of validity for predicting pilot training attrition and pilot training grades (Carretta, 2011). A smaller research literature has also shown some potential value of personality trait measures as predictors of pilot training performance, with meta-analyses showing Neuroticism and Introversion as significantly related to poor pilot training outcomes across studies (Campbell, Castaneda, & Pulos, 2012).

However, little is known regarding predictors of pilot performance post-training. To the extent that training is appropriately developed based on job requirements, one would expect the same predictors of pilot training performance would also predict pilot performance on-the-job. However, relatively few studies have examined the predictive validity of pilot aptitude and personality measures for post-training outcomes and these few studies focused on narrow rather than global or comprehensive performance criteria. For example, one study (Carretta, Perry, & Ree, 1996) showed cognitive ability (based on measures of divided attention, spatial reasoning, and working memory) predicted supervisor and peer ratings of situational awareness for 171 USAF F-15 pilots, after controlling for flight experience. No incremental validity was found for psychomotor or personality. Another study (King, Retzlaff, & Orme, 2000) examined the relations of the Big Five personality domains and their underlying facets to the occurrence of aircraft incidents and mishaps. Data were compared for pilots who were ($N = 28$) and were not ($N = 319$) judged to have played a causal role in the aircraft incident or mishap. Pilots who scored high

on self-assurance and devotion-to-duty were 3.75 and 2.39 times more likely to have committed pilot-error incidents/mishaps, respectively. Research based on mishaps are limited by their low base rate and potential difficulty in isolating contributing factors within the pilot's control. Beyond these limitations, we note that such research would fail to capture other key aspects of pilot job performance that are more contextual (see Borman & Motowidlo, 1993). For example, focusing on mishaps where a pilot could be judged to have played a causal role overlooks the broader performance expectations for pilots, particularly in the military. The highest performing pilots, for example, may not be just those who avoid mishaps when flying themselves, but also those who proactively speak up to help prevent potential errors by others or who readily volunteer for additional duties when needed to support the mission.

Hence, because little is known regarding the predictors of on-the-job performance even of manned aircraft pilots, a secondary purpose of this study was to determine if the current USAF pilot selection measures, which have been explicitly chosen based on their predictive relationship to pilot training outcomes, are also predictive of longer-term performance of pilots on the job.

To address the deficiency in post-training validation studies, the current study evaluated pre-accession aptitude and personality trait measures as predictors of manned and unmanned aircraft pilot performance on-the-job, as measured by supervisor and senior rater stratification on three years of Officer Performance Reports (OPRs).

Method

Participants

The samples consisted of 3,140 manned aircraft pilots and 330 RPA pilots whose job performance had been documented in an officer performance report (OPR) following completion of flight training. A subsample of 1,662 manned aircraft pilots and 146 RPA pilots who had three years of OPRs was analyzed separately. Participants completed the Air Force Officer Qualifying Test (AFOQT; Drasgow, Nye, Carretta, & Ree, 2010) pre-accession as a part of the officer commissioning and aircrew training selection process (United States Air Force, 2014).

Both the manned aircraft and RPA pilot samples were predominantly White (manned aircraft pilots: 90.73%; RPA pilots: 85.76%), and male (manned aircraft pilots: 93.87%, RPA pilots: 94.53%). Pilots

were typically first lieutenants (O-2) at the time of the first (post-training) OPR, and captains (O-3) at the time of the third OPR, but this varied due to factors such as duration of training, amount of time in casual status before and after training, and opportunities for below-the-zone (early) promotion.

Measures

Air Force Officer Qualifying Test (AFOQT).

The AFOQT Form S consists of 11 cognitive tests (Drasgow et al., 2010) and a measure of the Big Five personality domains (Self-Description Inventory+, Manley, 2011). The cognitive tests are combined into composite scores. Personnel decisions including qualification for officer commissioning programs and aircrew training are made, in part, on the basis of the cognitive composites (United States Air Force, 2014). Percentile scores (1-99) indicate one's approximate standing relative to the officer applicant pool (such that a score of 25, for example, indicates a score higher than roughly 25% of USAF officer applicants. Minimum qualification standards for officer commissioning are a Verbal composite score of 15 or greater and a Quantitative composite score of 10 or greater. Minimum qualification scores for pilot and RPA pilot training are a Pilot composite score of at least 25; additionally, a Combat System Officer composite score of at least 15 was required for manned aircraft and RPA pilots during the time period of this study. Table 1 shows which tests contribute to each composite. It should be noted that two of the AFOQT Form S cognitive tests, Rotated Blocks and Hidden Figures, did not contribute to any of the composites and are not included in this study.

Test-retest reliabilities for the composites were: Verbal ($\alpha = .89$), Quantitative ($\alpha = .84$), Academic Aptitude ($\alpha = .85$), Pilot ($\alpha = .84$), and Combat Systems Officer ($\alpha = .89$) (Parish, Morath, Lodato, & Stachowski, 2008). Internal consistency (α) reliabilities for the subtests ranged from .71 (General Science) to .90 (Instrument comprehension) (Morath, Parish, Lodato, Schwartz, & Weissmuller, 2008). Brief descriptions of the AFOQT cognitive tests are provided below.

Verbal tests. Verbal Analogies (VA) measures the ability to reason and determine relationships between words. Word Knowledge (WK) assesses verbal comprehension involving the ability to understand written language through the use of synonyms.

Table 1
Composition of AFOQT Aptitude Composites

Test	Reliability	V	Q	AA	P	CSO
Verbal Analogies	.74	X		X		X
Arithmetic Reasoning	.86		X	X	X	X
Word Knowledge	.87	X		X		
Math Knowledge	.85		X	X	X	X
Instrument Comprehension	.90				X	
Block Counting	.82					X
Table Reading	.89				X	X
Aviation Information	.81				X	
General Science	.71					X

Note. The composites are Verbal (V), Quantitative (Q), Academic Aptitude (AA), Pilot (P), and Combat Systems Officer (CSO).

Quantitative tests. Arithmetic Reasoning (AR) provides a measure of the ability to understand arithmetic relations expressed as word problems. Math Knowledge (MK) measures the ability to use mathematical terms, formulas, and relations.

Spatial ability. Block Counting (BC) assesses spatial ability through the analysis of three-dimensional representations of a set of blocks.

Perceptual speed. Table Reading (TR) measures the ability to quickly and accurately extract information from a table at a given set of X and Y coordinates.

Aircrew tests. Instrument Comprehension (IC) assesses the ability to determine the attitude of an aircraft from illustrations of flight instruments. Aviation Information (AI) evaluates knowledge of general aviation terms, concepts, and principles.

Science knowledge. General Science (GS) assesses knowledge and understanding of scientific terms, concepts, instruments, and principles.

Self-Description Inventory (SDI+). The SDI+ (Manley, 2011) is a 220-item trait-based measure that assesses the Big Five domains of neuroticism, extraversion, openness, agreeableness, and conscientiousness. It also includes a measure of Machiavellianism. It is administered as an AFOQT subtest. Unlike the AFOQT composites, on which minimum scores are required for commissioning and pilot training, there are no required minimum SDI+ scores. Internal consistency reliabilities for the Big Five domain scores ranged from .89 to .97 (.75 for Machiavellianism) for a sample of about 60,000 officer applicants (Manley, 2011).

Job performance criteria. Job performance criteria were derived from annual Officer Performance Reports (OPR). See Figure 1 for a de-identified

example. OPRs are generated annually to summarize an officer’s performance during the previous year. They are used to provide job performance feedback during annual supervisor reviews and are reviewed by promotion boards to make promotion decisions later in an officer’s career. The first truly competitive promotion decisions (about 80% selection rate) do not occur until O-4 (Major) promotion eligibility after approximately 10 years of Service (Hosek, Tiemeyer, Kilburn, Strong, & Ray, 2001). OPRs include a concise summary of career activities and their impact on mission performance. An important section of the OPR is narrative text provided by direct supervisors and senior raters (Colonel or higher). These two sections give raters an opportunity to make statements that distinguish top performers from others. Although OPRs provide only an overall assessment of performance, raters are directed to consider the performance factors of: communications skills, fitness, job knowledge, judgment and decision-making, leadership skills, organizational skills, and professional qualities.

There are strong organizational norms indicating the expectation that raters will numerically stratify (or, in USAF vernacular, “rack-and-stack”) high-performing subordinates (and only high-performing subordinates) as a means to inform later promotion decisions. Both official Air Force websites (see <http://www.afrc.af.mil/News/ArticleDisplay/tabid/267/Article/587530/air-force-reserve-epropref-writing-guide.aspx>) and unofficial commercial websites (e.g., afmentor.com) contain primers on writing effective OPRs, and these resources emphasize the importance of stratification statements as a message to senior officers who will review the records during later promotion boards. Official guidance (AFI 36-2406, section 1.12) provides specific examples of

authorized and prohibited stratification statements. This guidance directs supervisors to base OPR comments on “overall performance and performance-based potential as compared to others in the same grade known by the evaluator” (section 3.1.8.6), and clarifies that “evaluators can only stratify personnel within the confines of their direct rating chain and/or scope of responsibility” (section 1.12.1.3). Statements such as “top 5% officer,” for example, are identified as prohibited because the evaluator would not have first-hand knowledge of all Air Force officers; in contrast, guidance authorizes commanders to use stratification statements such as “#1/4 Lts in the squadron.” Because numeric stratification is not required, and rare cases of truly unacceptable

performance can be handled in separate formal processes (including letters of reprimand and referral reports), organizational norms dictate that only above-average stratification be included in OPRs. Hence while a supervisor may identify one subordinate as #1 of 5 (i.e., best-performing of the five individuals he supervises in similar positions) on that individual’s OPR, it is relatively uncommon to explicitly identify an average subordinate (#3 of 5) as such on an OPR, and unheard of to identify below average (e.g., #4 or 5 of 5). Across all of the pilots in this study, below average performance was never stratified, such that inclusion of a stratification statement consistently indicated average performance or higher.

III. PERFORMANCE FACTORS		DOES NOT MEET STANDARDS	MEETS STANDARDS
Job Knowledge, Leadership Skills (to include Promoting a Healthy Organizational Climate), Professional Qualities, Organizational Skills, Judgment and Decisions, Communication Skills (see reverse if marked Does Not Meet Standards)		<input type="checkbox"/>	<input checked="" type="checkbox"/>
IV. RATER OVERALL ASSESSMENT (Limit text to 6 lines)			
- Combat proven wingman: 97 combat sorties, 415 combat hrs, 30 XYZs, 27 ABCs in 6 months--100% on all tasks - Skilled attack pilot: precision delivery of 9 XYZs & 700 rnds 30mm in close prox of troops--saved coalition lives! - Volunteered to assist with critical engine maintenance: saved 15+ man hours--five A-10Cs combat ready in min t - Expertly led mission planning cell; supervised 3x CGOs/managed \$350K+ equip/50+ pubs--sq deployment ready - Excelled in maritime ops verification: tackled complex Navy/AF integration issues--invaluable to critical unit - My #1/7 CGOs: excellent aviator and combat warrior; rdy for flt lead upgrade now-- send to PDE soonest!			
Last performance feedback was accomplished on: _____ (IAW AFI 36-2406) (if not accomplished, state the reason)			
NAME, GRADE, BR OF SVC, ORGN, COMMAND & LOCATION		DUTY TITLE	DATE
		SSN	SIGNATURE 
V. ADDITIONAL RATER OVERALL ASSESSMENT (Limit text to 4 lines)		<input checked="" type="checkbox"/> CONCUR	<input type="checkbox"/> NON-CONCUR
- Special Operations force multiplier: responded to 2 Special Ops requests for XY--2 high-value targets neutralized - Skilled attack pilot: expertly employed in 8-ship/96 sortie ABC--32 special ops XYZ fully ready for combat - Planned 080 XY missions: provided flights key info--1500+ joint tactical air requests/250 direct action msns filled - My #2/17 flight leads: superb performance w/ tremendous impact on cmbt readiness; upgrade to flight lead ASAP			

Figure 1. Officer Performance Report (OPR) direct supervisor assessment (top) and senior rater assessment (bottom) with examples of stratification statements.

Procedure

OPR narrative text was examined to identify numerical stratification statements indicative of job performance. For the OPRs where numeric stratification was present, stratification statements followed one of two basic text patterns: 1) “Top N%” (e.g., in the “Top 10%” of company grade officers, lieutenants, instructor pilots, etc.), and 2) “#X of Y,” or “#X/Y” (e.g., “my #1 of 23” or #1/23). The highlighted texts in Figure 1 are examples of stratification statements. OPR statements for each

individual were coded for presence or absence of a stratification statement. Thus, for the analyses involving the first OPR, scores could be 0 or 1 (absence/presence of a stratification statement). For the analyses involving pilots/RPA pilots with three OPRs, stratifications were coded as a score ranging from 0 to 3, such that 0 indicated no stratification on any of the first three years of OPRs, and 3 indicated stratification on all of the first three years of OPRs.

Correlations were computed to examine the relations between AFOQT cognitive and SDI+ personality scores and OPR stratification scores. A

.05 Type I error rate and a directional test were used. Observed correlations were corrected for multivariate range restriction (Lawley, 1943) to provide a better estimate of the test score validities. Correlations involving the first OPR also were corrected for dichotomization (Cohen, 1983) of the OPR stratification score. The reference group for the range restriction correction consisted of commissioned USAF officers with AFOQT Form S scores. Correlations were not corrected for criterion attenuation, as an inter-rater reliability estimate of raters' numerical stratifications was not available.

Results

First Post-Training OPR

The first post-training OPR stratification rate was higher for pilots (44.2%) than for RPA pilots (27.5%)

$[t(3,468) = 5.84, p < .01]$. While unlikely the full source of the difference in stratification rates, the mean AFOQT Pilot score for manned aircraft pilots ($M = 78.61, SD = 18.68$) was higher than for RPA pilots ($M = 74.43, SD = 20.16$) [$t(3,468) = 3.83, p < .01$]. Consistent with organizational norms, stratification statements (when present) indicated a ranking that was at least average relative to an identified peer comparison group. When stratifications were present, the median stratification from supervisors placed the (RPA or manned aircraft pilot) ratee at the 80th percentile relative to an identified peer comparison group (e.g., #1 of 5 or #2 of 10); the modal stratification was at the 67th percentile (e.g., #1 of 3, or #2 of 6); the minimum stratification was at the 50th percentile (e.g., #1 of 2 or #2 of 4).

Cognitive ability. Table 2 summarizes the correlations between the AFOQT and SDI+ scores and stratification on the first post-training OPR separately for manned aircraft and RPA pilots. The Pilot composite and all of the subtests which

Table 2

Observed and Corrected Correlations between AFOQT Aptitude and Trait Measures and Stratification on First OPR (Presence/ Absence of Stratification from Supervisor and/or Senior Rater, Full Sample)

Score	Pilots ($N = 3,140$)			RPA Pilots ($N = 330$)		
	r	r_c	r_{c2}	r	r_c	r_{c2}
Pilot	.098**	.138	.175	.183**	.238	.332
VA	.009	.009	.011	-.021	-.023	-.032
AR	.055**	.062	.079	.093	.098	.137
WK	.014	.015	.019	.076	.079	.110
MK	.043*	.049	.062	.029	.033	.046
IC	.077**	.112	.142	.062	.087	.121
BC	.063**	.069	.087	.094	.102	.142
TR	.083**	.087	.110	.164**	.161	.224
AI	.060**	.073	.093	.216**	.247	.344
GS	.030	.035	.044	.094	.110	.153
SDI-A	.055**	.054	.068	.107	.114	.159
SDI-N	-.049**	-.051	-.065	-.141**	-.151	-.211
SDI-E	-.072**	-.073	-.093	-.062	-.065	-.091
SDI-C	.080**	.079	.100	.136*	.147	.205
SDI-O	-.004	-.004	-.005	.012	.012	.017
SDI-M	-.020	-.021	-.027	.002	.002	.003

Note. The columns labeled " r ," " r_c ," and " r_{c2} " are for the observed (uncorrected) correlations, correlations corrected for range restriction, and correlations corrected for both range restriction and dichotomization of the criterion. No significance tests were computed for the corrected correlations.

* $p \leq .05$; ** $p \leq .01$

contribute to it (AR, MK, IC, TR and AI) were significant predictors of manned aircraft pilot stratification. This included quantitative (MK and AR), perceptual speed (TR), and aviation (IC and AI). The verbal (VA and WK) and science knowledge (GS) subtests showed no significant relationships to OPR stratification. The magnitude of the relationships between the AFOQT Pilot composite and subtests and OPR stratification were generally similar, if not higher, for RPA pilots as compared to manned aircraft pilots. However, due to the smaller RPA pilot sample size, only the Pilot composite and the aviation knowledge (AI) and perceptual speed (TR) subtests were significantly related to stratification on the first post-training OPR. Applying the Fisher *r*-to-*z* transformation showed that the correlation between Aviation Information scores and OPR stratification was stronger for RPA pilots (uncorrected: .22; corrected: .32) than for manned aircraft pilots (uncorrected: .06; corrected: .12) [*z* = 2.82 for uncorrected correlations, *p* < .01]. For other cognitive scores, differences in predictive validity between manned aircraft and RPA pilots were not statistically significant.

Personality. Neuroticism (-) and Conscientiousness (+) were significantly related to OPR stratification and Openness and Machiavellianism had the lowest validities for both groups. Although the magnitude of the validities for Agreeableness (+) and Extraversion (-) were similar for both groups, the correlations were significant only for manned aircraft pilots due to its larger sample. While the uncorrected relationships between Neuroticism and OPR stratification were not significantly higher for RPA than for manned aircraft pilots (uncorrected *r*s = -.14 and -.05, respectively; *z* = 1.56, *p* = .06, two-tailed), the difference in magnitude for the range-restriction corrected validity coefficients was suggestive that the predictive validity of Neuroticism may be substantively higher for RPA (corrected *r* = -.19) than for manned aircraft (corrected *r* = .06) pilots. Differences in predictive validity between manned aircraft and RPA pilots were not statistically significant for the other SDI+ personality scores.

Table 3
Observed and Corrected Correlations between AFOQT Aptitude and Trait Measures and Number of Stratifications on First Three OPRs.

Score	Pilot (<i>N</i> = 1,662)		RPA Pilot (<i>N</i> = 146)	
	<i>r</i>	<i>r_c</i>	<i>r</i>	<i>r_c</i>
Pilot	.105**	.168	.132	.200
VA	.016	.018	.015	.019
AR	.069**	.080	-.045	-.018
WK	-.021	.003	.008	.009
MK	.066**	.074	-.089	-.072
IC	.056*	.090	.095	.150
BC	.054*	.061	-.024	.008
TR	.131**	.139	.160	.155
AI	.039	.053	.082	.109
GS	.025	.029	-.086	-.059
SDI-A	.091**	.090	.121	.127
SDI-N	-.045	-.047	-.165*	-.173
SDI-E	-.076**	-.079	-.155	-.170
SDI-C	.110**	.108	.263**	.291
SDI-O	-.039	-.040	-.135	-.146
SDI-M	-.043	-.044	-.083	-.084

Note. The columns labeled “*r*,” and “*r_c*” and are for the observed (uncorrected) correlations and correlations corrected for range restriction. No significance tests were computed for the corrected correlations.

p* ≤ .05; *p* ≤ .01

Pilots/RPA Pilots with Three OPRs

The mean number of OPRs with stratification statements (across a pilots' first three years post-training, 0-3) was higher for manned aircraft pilots ($M = 1.49$, $SD = 1.01$) than for RPA pilots ($M = 1.32$, $SD = 0.90$) [$t(1,806) = 1.84$, $p < .05$]. However, the magnitude of the difference was smaller than was observed for the first OPR stratification score.

Cognitive ability. Table 3 summarizes the correlations between the AFOQT and SDI+ scores and stratification score for the first three OPRs for manned aircraft and RPA pilots. Correlational results for the cognitive tests were similar to those for the analyses involving the first OPR. For manned aircraft pilots the Pilot composite and 4 of its 5 subtests (AR, MK, IC, and TR) were significantly related to the OPR stratification score. The two verbal subtests (VA and WK) showed the lowest correlations with OPR score. While the magnitude of correlations was generally similar for manned and unmanned aircraft pilots, the correlations were not statistically significant for RPA pilots due to smaller sample size.

Personality. For manned aircraft pilots, results were similar to those observed for the first OPR stratification criterion. Agreeableness (+), Extraversion (-), and Conscientiousness (+) were significantly related to stratification score and Openness and Machiavellianism were not. Contrary to the first OPR results, Neuroticism was not significantly related to OPR score.

The results for RPA pilots were similar to those for the first OPR stratification criterion. That is, Neuroticism (-) and Conscientiousness (+) were significantly related to the OPR score. Although there was the same trend for Neuroticism (-) to more strongly predict performance for RPA ($r = -.17$) than for manned aircraft pilots ($r = -.05$), the correlations did not significantly differ given the small RPA sample size with three years of OPRs available. Uncorrected correlations between Conscientiousness (+) and OPR stratification were not significantly higher for RPA (uncorrected $r = .26$) than for manned aircraft (uncorrected $r = .11$) pilots [$z = 1.79$, $p =$

$.07$, two-tailed], although the pattern of range-restriction corrected validity coefficients suggested substantively higher validity of Conscientiousness for predicting stratifications among RPA than manned aircraft pilots (corrected $r = .29$ for RPA as compared to corrected $r = .11$ for manned aircraft pilots).

Discussion

Overall, the cognitive aptitude, knowledge, and personality traits that are predictive of USAF manned aircraft pilot job performance also are predictive of RPA pilot job performance. Consistent with current USAF pilot selection requirements, the subtests that contribute to the AFOQT Pilot composite -- including quantitative, perceptual speed, and aviation knowledge measures -- were significant predictors of early career job performance for both manned aircraft and RPA pilots. After corrections for range restriction and criterion dichotomization, but not criterion attenuation, the AFOQT-Pilot composite was correlated .18 with first-year job performance for manned aircraft pilots and .33 for RPA pilots. These results are important because they demonstrate that the current entry standards, which were developed to predict manned aircraft training completion, also predict early on-the-job post-training performance for both manned aircraft and RPA pilots.

Although previous studies have shown that the cognitive aptitude and trait predictors of manned aircraft pilot training success are also effective in predicting USAF RPA pilot training outcomes (Carretta, 2013; Rose et al., 2014), one possibility was that these relationships were an artifact of the decision by the USAF to explicitly design the RPA training curriculum to parallel that for manned aircraft pilots, including requiring manned aircraft flight in initial flight screening, and simulated flight of the same aircraft used in manned aircraft pilot training in later training phases (RPA Instrument Qualification). Results from the current study demonstrate that the similar results for manned aircraft and RPA pilots are not an artifact of training design; the same predictive relationships were observed for manned aircraft and RPA pilots on the job.

While the magnitude of correlations was not strong (Cohen, 1988; U. S. Department of Labor, 1999) for either manned aircraft or RPA pilots (all r s $< .35$ after correction for range restriction and dichotomization), the reported validities should be considered as very conservative estimates of the true relationship between pre-accession measures and job performance as the data were not corrected for criterion attenuation. In USAF pilot training, performance of each flight maneuver is carefully tracked and evaluated in each daily sortie, and ratings are aggregated and standardized relative to individuals in the same class and rated by the same instructors (Carretta et al., 2014). In contrast, because on-the-job performance is much more difficult to assess in a standardized, reliable manner (see Viswesvaran, Ones, & Schmidt, 1996), it would be extremely rare for any employee's performance to be as carefully observed and as directly, numerically evaluated as occurs in

USAF pilot training. Hence, because job performance ratings are likely to be much more subject to criterion attenuation than the aggregate instructor ratings of pilot training performance used in past studies (e.g., Carretta, 2013; Rose et al., 2014), a direct comparison of the magnitude of the reported validity coefficients (even after corrections for range restriction and criterion dichotomization) of pilot selection tests for predicting training vs. job performance is inappropriate.

Our use of an overall post-training, job performance metric as a criterion for validating pilot selection tests is, to our knowledge, a unique contribution to the literature. However, while archival stratification statements were found to be a reasonable proxy for pilots' job performance, we recognize that they are a highly imperfect indicator. We would have preferred to have supervisory job performance rankings available for all pilots (rather than purposely omitted for less stellar rates), and we would have preferred if (when stratification information was available) rankings were made relative to a standardized peer comparison group. However, military operations often make research-only data collection of highly-standardized, behaviorally-anchored, supervisory ratings of on-the-job, post-training performance infeasible, necessitating use of archival data. To the extent that future military researchers may have such strong organizational support as to allow for more standardized post-training performance ratings of pilots in an operational environment, we envy them.

Although a direct comparison of the magnitude of validity coefficients reported in the present study of on-the-job performance to the magnitude of validity coefficients reported based on training criteria is not appropriate, the pattern of significant trait and aptitude predictors is instructive. Where the pattern of significant findings based on training outcomes differed from those based on on-the-job performance as measured by OPR stratification, the results differed for both manned aircraft and RPA pilots. That is, previous USAF studies generally have failed to find consistent significant relationships between Big Five traits and manned aircraft (Carretta et al., 2014) or RPA (Carretta, 2013; Rose et al., 2014) pilot training outcomes, other than the finding that Openness to Experience is modestly negatively associated with pilot training completion/grades. In contrast, the current study found Conscientiousness (+) and Neuroticism (-) to significantly predict post-training officer stratification for both manned aircraft and RPA pilots. These results are generally consistent with meta-analyses across varied civilian and military occupations that show Emotional Stability and

Conscientiousness as useful predictors of job performance (e.g., Barrick & Mount, 1991). Results are also consistent with a small meta-analysis (Campbell, Castaneda, & Pulos, 2010) showing significant validity of self-report assessments of Emotional Stability as a predictor of manned aircraft pilot training outcomes, at least for studies in which training graduation versus self-elimination ("Drop on request") was the criterion.

Differential Validity for RPA Relative to Manned Aircraft Pilots

The most striking findings of this study suggest that pre-accession measures of aviation knowledge, neuroticism (i.e., emotional stability), and conscientiousness may potentially be stronger predictors of job performance for RPA pilots than for manned aircraft pilots. Some possible theoretical explanations for the stronger relationships for RPA pilots for each of these three pre-accession measures are provided in turn. As formal evaluation of the underlying mechanisms was beyond the scope of this paper, we would encourage future research to formally evaluate these (or other) potential explanations.

Aviation knowledge. Although the content of the aviation knowledge test has applicability to both manned aircraft and RPA (i.e., aerodynamics, meteorology, navigation, and general aircraft parts and terminology), none of the content is specific to RPA. Because the aviation knowledge test is administered pre-accession, before undergoing any required training, it may function in part as an indicator of general interest and motivation in aviation, such that individuals who are highly motivated for a pilot career would actively pursue opportunities to learn about aviation and dedicate time to self-study in preparation for the test. One reason that pre-accession aviation knowledge may better predict performance for RPA than manned aircraft pilots is that entering the RPA career field with high interest and motivation may be particularly important for sustaining high levels of performance given the additional stressors and unpleasant job aspects faced by RPA pilots (Paullin, Ingerick, Trippe, & Wasko, 2011; U.S. Government Accountability Office, 2014). Whereas the opportunity to physically experience flight (e.g., aircraft vibration, pulling "gs") may sustain motivation in manned aircraft flight, performing duties from windowless rooms, for example, may dampen initial levels of motivation for RPA pilots.

Neuroticism. Neuroticism (i.e., emotional stability) may be a stronger predictor of job performance for RPA pilots than for manned aircraft pilots because, in some ways, RPA pilots are asked to

perform a more stressful job. Potential stressors for RPA as compared to USAF manned aircraft pilots include greater responsibility as a junior officer, challenging work-life balance issues associated with participating in combat operations while living at home, constantly rotating shifts, and more direct and extensive on-the-job monitoring. Previous research supports the notion that the extent of job stressors may moderate the relationship between emotional stability and job performance. For example, Spector, Jex, and Chen (1995) found that employees low in emotional stability were less likely to pursue or persist in jobs that entail potential stressors such as high levels of task variety, feedback, and complexity. Consistent with this finding, compared to individuals with high emotional stability, those with low emotional stability rely on ineffective coping strategies at higher levels of stress (Suls & Martin, 2005), are more reactive to aversive work events (Bolger & Zuckerman, 1995; Dunkley, Mandel, & Ma, 2014), and generally perceive higher levels of job strain and lower levels of control (Tornroos et al., 2013).

Conscientiousness. Frequent observation by leadership, in addition to serving as a stressor for RPA pilots, may also account for a stronger relationship between Conscientiousness and job performance for RPA relative to manned aircraft pilots. Without the geographic separation of pilots in manned aircraft, leadership can more easily directly monitor RPA pilots' actions in real-time. By more extensively and consistently monitoring RPA pilots' daily performance, supervisors may be more likely to notice unconscientious behavior than they would if they were observing performance from afar. Hence, potentially lower validity of conscientiousness for manned aircraft pilots may reflect that they are more likely than RPA pilots to be assessed based on their maximum rather than typical performance (see Klehe & Anderson, 2007).

Implications for RPA Pilot Selection and Recruiting

While the results of this study generally suggest that the same selection methods predictive of performance as a manned aircraft pilot are predictive of performance as an RPA pilot, the process for recruiting RPA candidates may need to differ substantially. Even the most valid selection methods will only be effective to the extent that the Air Force has more RPA applicants (volunteers for RPA training) than it has seats to fill. The Air Force is planning to increase production of active-duty URT graduates from 180 per year to approximately 300 per year (Schanz, 2015). While the Air Force has taken

steps to increase incentive pay and retention bonuses for RPA pilots, without additional recruitment efforts, the demand for qualified RPA applicants may exceed supply (Schanz, 2015). Although a recent policy change requires individuals who volunteer for manned aircraft piloting to also volunteer for RPA, the manned aircraft pilot career remains the first choice for most aircrew applicants.

In addition to generating applicants, the Air Force must address the difficult task of recruiting individuals with a high level of interest and enthusiasm for the RPA career field to ensure that those selected to attend URT are as highly motivated as those selected to attend manned aircraft training. Even without the thrill of "real" flying as a recruiting tool, the task may not be impossible. One tool that may help recruit applicants with an interest in the RPA career field is a person-job match assessment based on individual preferences for work context factors in the RPA environment (Work Interest Inventory or WII, Paullin et al., 2011). Initial results have shown RPA pilot incumbents' and trainees' WII responses to be related to job satisfaction and intent to stay in the Air Force after their initial service commitment (Barron, Rose, & Burns, 2015).

While the negative aspects of the RPA environment (e.g., windowless rooms, rotating shifts) likely already are well-known to pilot applicants, feedback from the WII may improve potential RPA candidates' awareness of positive aspects of the RPA work environment. Although manned aircraft likely will continue to be the first choice for the majority of those interested in a military aviation career, there are several factors that may increase the attractiveness of RPAs. First, as noted earlier, RPA career opportunities in both the military and civilian sectors are expected to grow as technology matures, new applications are explored, and regulations are developed to integrate USA/RPA in the national airspace. Further, we suspect that many high-aptitude cadets may be more likely to seek out RPA careers if they were fully aware of the opportunities RPA offers for greater responsibility (serving in a leadership role as an aircraft commander) early in their career and greater potential for flying high-profile combat missions.

References

Aerospace Industries Association (2013). *Unmanned aircraft systems: Perceptions and potential*. Arlington, VA: Author.

- Barrick, M. R., & Mount, M. K. (1991). The Big Five personality dimensions and job performance. *Personnel Psychology, 44*, 1-26.
- Barron, L. G., Carretta, T. R., & Bonto-Kane, M. V. (in press). Relations of aviator personality traits to aviator performance: It depends on the criterion.
- Barron, L. G., Rose, M. R., & Burns, G. N. (2015). *Initial validation of the work interest inventory for remotely-piloted aircraft (RPA) career fields*. Paper presented at the 18th International Symposium on Aviation Psychology, Dayton, OH.
- Bolger, N., & Zuckerman, A. (1995). A framework for studying personality in the stress process. *Journal of Personality and Social Psychology, 69*, 890-902.
- Borman, W. C., & Motowidlo, S. M. (1993). Expanding the criterion domain to include elements of contextual performance. In N. Schmitt and W. C. Borman (Eds.), *Personnel selection in organizations* (pp. 71-98). San Francisco: Jossey-Bass.
- Campbell, J. S., Castaneda, M., & Pulos, S. (2010). Meta-analysis of personality assessments as predictors of military aviation training success. *International Journal of Aviation Psychology, 20*, 92-109.
- 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*, 47-53.
- Carretta, T. R., Perry, D. C., Jr., & Ree, M. J. (1996). Prediction of situational awareness in F-15 pilots. *International Journal of Aviation Psychology, 6*, 21-41.
- Carretta, T. R., & Ree, M. J. (2003). Pilot selection methods. In B. H. Kantowitz (Series Ed.) & P. S. Tsang & M. A. Vidulich (Vol. Eds.), *Human factors in transportation: Principles and practices of aviation psychology* (pp. 357-396). Mahwah, NJ: Erlbaum.
- Carretta, T. R., Teachout, M. S., Ree, M. J., Barto, E. L., King, R. E., & Michaels, C. F. (2014). Consistency of the relations of cognitive ability and personality traits to pilot training performance. *International Journal of Aviation Psychology, 24*, 247-264.
- 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.)*. Mahwah, NJ: Erlbaum.
- Damos, D. L. (2011). *A summary of the technical pilot selection literature*, AFCAPS-FR-2011-0009. Randolph AFB, TX: Air Force Personnel Center, Strategic Research and Assessment Branch.
- Dillingham, G. L. (2012). *Unmanned aircraft systems: Use in the national airspace system and the role of the Department of Homeland Security* (Report No. GAO-12-889T). Washington, DC: Government Accountability Office.
- Drasgow, F., Nye, C. D., Carretta, T. R., & Ree, M. J. (2010). Factor structure of the Air Force Officer Qualifying Test form S: Analysis and comparison with previous forms. *Military Psychology, 22*, 68-85.
- Dunkley, D. M., Mandel, T., & Ma, D. (2014). Perfectionism, neuroticism, and daily stress reactivity and coping effectiveness 6 months and 3 years later. *Journal of Counseling Psychology, 61*, 616-633.
- Everstine, B. (2015). DOD plans 50 percent increase in RPA CAPS by 2019. *Air Force Magazine*. Retrieved from <http://airforcemag.com/drarchive/pages/2015/august%202015/august%2018%202015/dod-plans-50-percent-increase-in-rpa-caps-by-2019-.aspx?signon=false>
- Gertler, J. (2012). *US unmanned aerial systems* (Report No. R42136). Washington, DC: Congressional Research Service.
- Greene Sands, R.R., & Dove-Steinkamp, M.L. (2015). *Small Wars Journal*. Retrieved from <http://smallwarsjournal.com/jrml/art/finding-meaning-inside-the-box-understanding-rpa-crew-resilience>
- Hosek, S. D., Tiemeyer, P. M., Kilburn, M. R., Strong, S. D., & Ray, R. (2001). *Minority and gender differences in officer career progression*, MR-1184-OSD. Santa Monica, CA: RAND Corporation. Retrieved from http://www.rand.org/pubs/monograph_reports/MR1184
- Hunter, D.R. & Burke, E.F. (1994). Predicting aircraft pilot-training success: A meta-analysis of published research. *International Journal of Aviation Psychology, 4*, 297-313.
- Ingerick, M. (2010). *Identifying leader talent: Alternative predictors for U.S. Air Force junior officer selection and assessment*, AFCAPS-FR-2010-0020. Randolph AFB, TX: Air Force Personnel Center, Strategic Research and Assessment Branch.

- Jenkins, D. & Vasigh, B. (2013). *The economic impact of unmanned aircraft systems integration in the United States*. Arlington, VA: Association for Unmanned Vehicle Systems, International.
- King, R. E., Retzlaff, P. D., & Orme, D. R. (2000). *A comparison of US Air Force pilot psychological baseline information to safety outcomes*, AFSC-TR-2001-0001. Kirtland AFB, NM: Policy Research, and Technology Division, Air Force Safety Center.
- Klehe, U., & Anderson, N. (2007). Working hard and working smart: Motivation and ability during typical and maximum performance. *Journal of Applied Psychology*, 92, 978-992.
- Lawley, D. N. (1943). A note on Karl Pearson's selection formulae. *Proceedings of the Royal Society of Edinburgh, Section A*, 62, Part 1, 28-30.
- Lentz, E., Horgen, K. E., Borman, W. C., Dullaghan, T. R., Smith, T., Schwartz, K. L., & Weissmuller, J. J. (2009). *Air Force officership survey volume I: Survey development and analyses*, AFCAPS-FR-2010-0010. Randolph AFB, TX: Air Force Personnel Center, Strategic Research and Assessment Branch.
- Manley, G. G. (2011). *Development of domain and facet level scales for the Self-Description Inventory*, ALCAPS-TR-2001-0007. Randolph AFB, TX: Air Force Personnel Center, Strategic Research and Assessment Branch.
- Martinussen, M. (1996). Psychological measures as predictors of pilot performance: A meta-analysis. *International Journal of Aviation Psychology*, 6, 1-20.
- Morath, R., Parish, C., Lodato, M., Schwartz, K., & Weissmuller, J. (2008). *AFOQT form S effectiveness analyses: Subtest-level analyses and report*, AFCAPS-FR-2010-0011. Randolph AFB, TX: Air Force Personnel Center, Strategic Research and Assessment Branch.
- Parish, C., Morath, R., Lodato, M., & Stachowski, A. (2008). *AFOQT form S Effectiveness: Composite-level analyses and report*. Fairfax, VA: ICF International.
- Paullin, C., Ingerick, M., Trippe, D. M., & Wasko, L. (2011). *Identifying best bet entry-level selection measures for US Air Force remotely piloted aircraft (RPA) and sensor operator occupations*, FR-11-64. Randolph AFB, TX: Air Force Personnel Center, Strategic Research and Assessment Branch.
- Rose, M. R., Barron, L. G., Carretta, T. R., Arnold, R. D., & Howse, W. R. (2014). Early identification of unmanned aircraft pilots using measures of personality and aptitude. *International Journal of Aviation Psychology*, 24, 36-52.
- Roughton, Randy (2015). Daily deployed: RPA crew members prepare each day they arrive for duty. *Airman*. Retrieved from <http://airman.dodlive.mil/2015/07/daily-deployed/>
- Schanz, M.V. (2015). Combat air patrol relief. *Air Force Magazine*, 98, 16.
- Spector, P. E., Jex, S. M., & Chen, P. Y. (1995). Relations of incumbent affect-related personality traits with incumbent and objective measures of characteristics of jobs. *Journal of Organizational Behavior*, 16, 59-65.
- Suls, J., & Martin, R. (2005). The daily life of the garden-variety neurotic: Reactivity, stressor exposure, mood spillover, and maladaptive coping. *Journal of Personality*, 73, 1-25.
- Toornroos, M., Hintsanen, M., Hintsanen, T., Jokela, M., Pulkki-Raback, L., Hutri-Kahonen, N., & Keltikangas-Jarvinen, L. (2013). Associations between five-factor model traits and perceived job strain: A population-based study. *Journal of Occupational Health Psychology*, 18, 492-500.
- Tvaryanas, A.P. (2006). Human systems integration in remotely piloted aircraft operations. *Aviation, Space, and Environmental Medicine*, 77, 1278-1282.
- United States Air Force (2014). *Air Force Strategic Personnel Assessment Program* (AFI-36-2136). Washington, DC, Author.
- U.S. Department of Labor (1999). *Testing and assessment: An employer's guide to good practices*. Retrieved from <http://wdr.doleta.gov/opr/fulltext/document.cfm?docn=6032>
- U.S. Government Accountability Office (2014). Air Force actions needed to strengthen management of unmanned aerial system pilots. Retrieved from <http://www.gao.gov/assets/670/662467.pdf>