2015

The Role of Personnel Selection in Remotely Piloted Aircraft Human System Integration

Thomas R. Carretta

Raymond E. King

THE ROLE OF PERSONNEL SELECTION IN REMOTELY PILOTED AIRCRAFT HUMAN SYSTEM INTEGRATION

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

Raymond E. King
FAA Civil Aerospace Medical Institute
Oklahoma City, OK

Effective human-system integration (HSI) incorporates several domains: manpower, personnel, and training, human factors, environment, safety, occupational health, habitability, survivability, logistics, intelligence, mobility, and command and control. These domains are interdependent and must be considered in terms of their interrelationships. Human factors engineers typically focus on system design with little attention to the skills, abilities, and other characteristics needed by the human operator. Personnel selection is seldom considered during the HSI process. Complex systems require careful selection of the individuals who will interact with the system. Selection is a two-stage process: *Select-in* procedures determine who has the aptitude to profit from a training program and, thus, represents the best investment. *Select-out* procedures focus on medical qualification and disqualification. Generally, less expensive screenings methods should be used first and more expensive filters used in later stages of the selection process. Personnel selection has a vital role to play in human-system integration.

Achieving high levels of effectiveness for complex systems such as remotely piloted aircraft (RPA) cannot be done solely through technological advances. Systems such as RPA consist of hardware, software, and human personnel which must effectively work together to achieve organizational objectives. Human-systems integration (HSI) is a comprehensive management and technical approach to address the role of human operators in system development and acquisition. HSI incorporates several domains including manpower, personnel, and training, human factors, environment, safety, occupational health, habitability, survivability, logistics, intelligence, mobility, and command and control (United States Air Force, 2014). These domains are interdependent and their interrelationships must be considered. HSI must be considered early in the system development and acquisition process to be effective. It is difficult and costly, if not impossible, to “fix” a poorly designed complex system after it has been built and implemented. This paper discussed the role of personnel measurement and selection for HSI, the development of Undergraduate RPA training (URT) selection standards, other important considerations in personnel selection, and expected changes in selection requirements as RPAs evolve.

Role of Personnel Measurement and Selection for HSI

Those responsible for human-system integration should be aware of the relations between selection, training, and human-system design and how they interact to affect overall system effectiveness. Poor personnel measurement and selection will result in higher training attrition and training costs, increased human-system integration costs, and lower levels of job performance. Poor training will require higher quality applicants and improved human-systems design to mitigate its effects. Poor human factors (i.e., clumsy automation, operator-vehicle interface design) will increase operator cognitive demands and workload, resulting in increased selection and training requirements. Effective selection (Carretta & Ree, 2003) and training (Patrick, 2003; Smallwood & Fraser, 1995) methods and human-automation interaction (Paraduraman & Byrne, 2003) can help reduce life cycle costs and contribute to improving organizational effectiveness.

The Development of Undergraduate RPA Training (URT) Selection Standards

US Air Force (USAF) RPA Pilot Selection Methods

In the USAF, early efforts to field RPA systems focused on technology development. Personnel selection, training, and human-interface design were given little attention, as the RPA system manning approach was to retrain manned aircraft pilots to operate RPAs. Although this approach was mostly effective, as demand for the capabilities provided by RPAs increased, it became too costly and unsustainable. In 2009, an Undergraduate RPA training

(URT) program was established to train personnel with no prior flying experience to operate RPAs. URT curricula were developed and selection requirements based on those for manned aircraft pilot training were established.

URT selection methods involve both select-in-and select-out procedures and are very similar to those for manned aircraft pilot training. Aptitude testing and Medical Flight Screening are two important factors. Aptitude testing includes the Air Force Officer Qualifying Test (AFOQT; Drasgow, Nye, Carretta, & Ree, 2010), Test of Basic Aviation Skills (TBAS; Carretta, 2005), and Pilot Candidate Selection Method (PCSM; Carretta, 2011). Aptitude requirements for URT qualification are identical to those for manned aircraft pilot training. Medical Flight Screening (MFS) includes successful completion of a FAA Class III Medical Certificate and an USAF Flying Class IIU Medical Examination (United States Air Force, 2011), review of medical records, psychological testing, and an interview. Results from the MFS psychological testing and interview are not used as part of a select-out process with strict minimum qualifying scores. Rather, a licensed psychologist uses clinical judgment to assess the psychological disposition of URT applicants to determine whether there is an aeromedically disqualifying condition in accordance with Air Force guidelines (United States Air Force, 2011). Results of two recent USAF predictive validation studies (Carretta, 2013; Rose, Barron, Carretta, Arnold, & Howse, 2014) have demonstrated similar levels of validity for the AFOQT Pilot and PCSM composites to those observed for manned aircraft pilot training.

Results for studies examining the utility of personality for URT are less consistent (Chappelle, McDonald, Heaton, Thompson, & Hanes, 2012; Rose et al., 2014). Chappelle et al. examined the predictive validity of the AFOQT Pilot composite, Revised NEO Personality Inventory (NEO-PI-R; Costa & McCrae, 1992), and a neuropsychological battery, the MicroCog (Powell, Kaplan, Whitla, Weintraub, Catlin, & Funkenstein, 2004) versus URT completion. The best-weighted regression composite for predicting URT completion included the AFOQT Pilot composite, several NEO-PI-R scales, and the MicroCog Reaction Time subtest. Discriminant analyses showed that personality scores improved classification accuracy (identification of true positives and true negatives) beyond that provided by cognitive ability and prior flight time. Classification accuracy improved from 57.1% to 75.2% when personality scores were included; however, these results likely capitalized on chance given the large number of NEO-PI-R scales relative to the small sample size.

Rose et al. (2014) examined the extent to which scores from a Big Five measure of personality could improve prediction of URT completion and training grades beyond the AFOQT Pilot and PCSM composites. Regression analyses showed no incremental validity for personality scores when used in combination with the AFOQT Pilot or PCSM composite scores for predicting URT completion. However, the Openness score demonstrated small, but statistically significant incremental validity for predicting initial RPA qualification training grade.

RPA System Job/Task Analyses

Despite the predictive validity of current RPA pilot training selection methods, several studies have been conducted to determine whether there are any unique job-related skills, abilities, and other characteristics (SAOCs) not adequately measured by current selection methods (for a summary, see Carretta, Rose, & Bruskiewicz, in press; Williams, Carretta, Kirkendall, Barron, Stewart, & Rose, 2014). In the Williams et al. study, Air Force, Army, and Navy subject matter experts in personnel measurement, selection, and testing identified and assigned importance ratings to 115 SAOCS that appeared in one or more military RPA job/task analyses. Where available, psychometric data for existing DoD and US Military Service proprietary personnel selection and classification tests were examined to determine the extent to which the tests measure critical RPA SAOCs and to identify measurement gaps. Seventy-eight of the 115 SAOCs received an average rating of 3 (moderately important) or higher on a 5-point scale. Of these, 57 of 78 (73%) were judged to be measured by one or more existing military proprietary tests. It is interesting to note that many of the most important SAOCs involved personality (e.g., conscientiousness, stress management, dependability, vigilance, adaptability/flexibility, integrity, responsibility, self-discipline). Table 1 provides examples of the highest-rated cognitive, personality/temperament, and other characteristics. See Williams et al. for the complete list of SAOCs.

Williams et al. (2014) made several recommendations for RPA operator test battery content. As previously noted, most of the critical SAOCs were judged to be measured by existing proprietary DoD or US Military Service tests. They recommended that a program be established to increase the reliability and reduce the fakeability of military personality tests. They also recommended that new tests be developed to fill measurement gaps (e.g., oral comprehension, vigilance) and to improve experimental measures involving task prioritization/multi-tasking and work preferences (person-environment fit).

Table 1.

Examples of SAOCS Rated Most important for RPA Pilots

Cognitive	Personality/Temperament	Other
Task Prioritization	Conscientiousness	Time Sharing
Oral Comprehension	Stress Management/Tolerance	Control Precision
Spatial Orientation	Dependability	Occupational Interests/ Work Preferences, P-E Fit
Oral Expression	Vigilance (ability & personality)	
Attention to Detail	Adaptability/Flexibility	
Critical Thinking	Responsibility	
	Self-Discipline	

Other Important Considerations in Personnel Selection

The Criterion

Many researchers spend enormous amounts of effort to develop measures of critical SAOCs based on the results of job/task analyses. They then search for available, convenient, or easy to collect job performance criteria with little thought about the theoretical meaning or psychometric properties of the criteria. The same care used to develop personnel selection methods and predictors of job performance should go into the development of job performance criteria. Failure to consider the psychometric properties of the criterion (e.g., construct validity, dimensionality, discriminability, reliability) leads to incorrect decisions about the effectiveness of selection methods and their relation to job performance. Problems also are caused by inattention to relevancy, contamination, and deficiency of the criterion.

As with measures used for personnel selection, job performance criteria vary in the constructs they measure, content, and specificity. To the extent possible, the constructs assessed by the job performance criteria should match those measured by the selection measures. As we have discussed, RPA job/task analyses have identified several critical personality traits needed for success. However, predictive validation studies have shown relatively low validities for personality compared to cognitive and other measures. One reason for this finding may be the job performance criteria used in these studies do not capture constructs for which personality is important (e.g., effort, leadership, indicators of maladaptive or counterproductive behavior). McHenry, Hough, Toquam, Hanson, and Ashworth (1990) provided an example that demonstrates the importance of criterion specificity. McHenry et al. administered a large battery of measures including ability and personality/temperament to a sample of U. S. Army trainees. Multiple criteria were used to reflect different aspects of job performance. Cognitive tests were the best predictors of criteria reflecting technical job proficiency, while measures of personality/temperament were the best predictors of criteria reflecting effort and leadership.

Special Population Norms

The assessment of human characteristics is based on comparing an individual to a representative sample of the population. Certain segments of the population vary significantly from the general population. For example, groups may differ on level of academic achievement, physical fitness, job experience, specialized knowledge/training, or other factors related to occupational performance. Moreover, differences in personality across occupational groups such as sales personnel, pilots, and engineers may occur. Military aircrew personnel are a highly selected and distinguished occupational group. Competition for pilot training assignments is great with the result that those selected differ significantly from the general adult population on cognitive, personality, and other characteristics considered during the selection process. Carretta et al. (2014) reported cognitive and personality norms for large samples of USAF pilot trainees. They observed that the mean full-scale IQ score for this group (M = 120) was about 1.33 SDs above the normative adult population mean (M = 100, SD = 15)). A pilot with a mean full-scale IQ of 105 would be slightly above the normative adult population mean, but well below the mean for USAF pilot trainees.

Significant differences also have been observed for personality scores of USAF pilot trainees compared to adult population norms. The personality portion of the USAF Neuropsychiatrically Enhanced Flight Screening (King &

Flynn, 1995) program, the forerunner of MFS, was developed to compile special population norms. The battery has been composed of the 1) Armstrong Laboratory Aviation Personality Survey (Retzlaff, Callister, & King, 1997) and 2) NEO Personality Inventory-Revised (Costa & McCrae, 1989). The ALAPS measures personality, psychopathology, and crew interaction, while the NEO-PI-R measures the Big Five domains and facets of normal personality. Figure 1 illustrates the number of standard deviations USAF pilot normative means are above or below those for the adult general population. Similar specialized norms are not presented for the ALAPS because it was normed on a USAF student pilot sample.

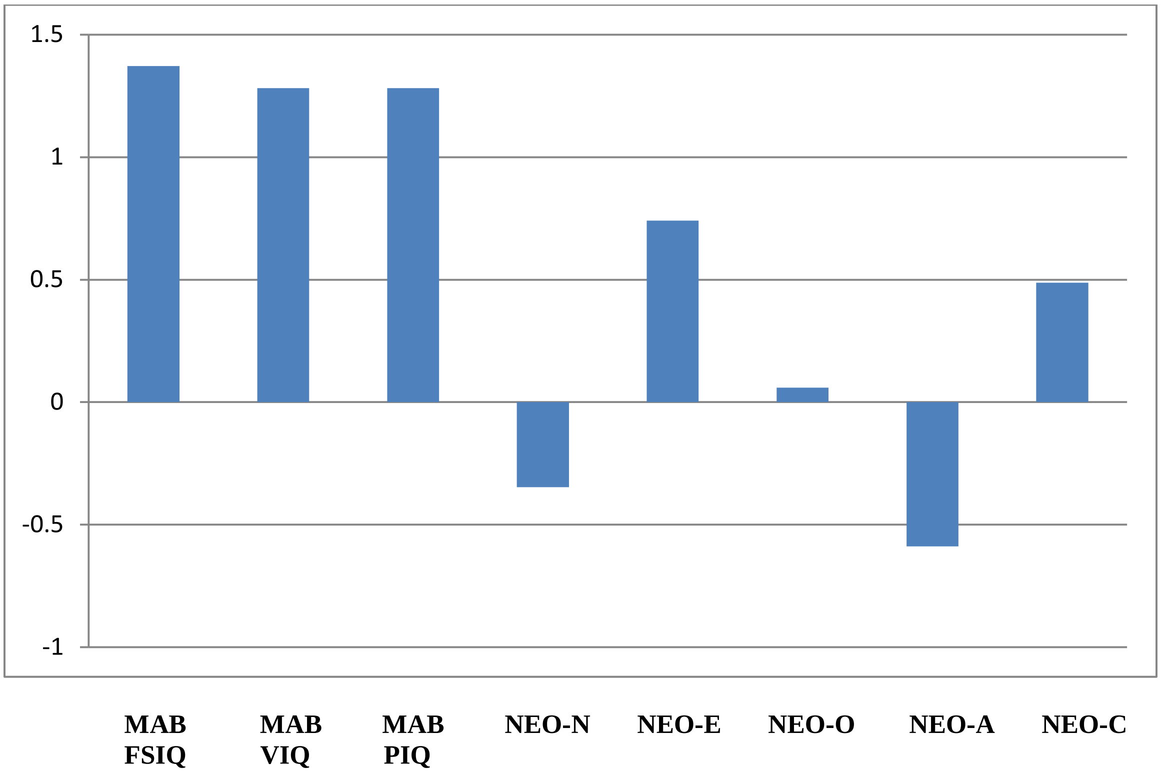

Figure 1. US Air Force pilot trainee norms versus the adult general population. The bars indicate the number of standard deviations the US Air Force pilot trainee means are above or below the adult population means. The scores are the Multidimensional Aptitude Battery (MAB) Full-Scale IQ (FSIQ), Verbal IQ (VIQ), and Performance IQ (PIQ) and the NEO-PI-R Neuroticism (N), Extraversion (E), Openness (O), Agreeableness (A), and Conscientiousness (C) scores.

To date, over 26,000 USAF student pilots have been administered some combination of these psychological tests. King, Barto, Ree, and Teachout (2011) presented a compendium of specialized USAF personality testing norms that can be used with military pilots and, cautiously, with applicants for civil airlines. This report includes profile sheets tailored specifically with these norms. A perusal of these norms demonstrates that USAF pilots differ from the general population on commercially published test norms. For example, this population has a mean *Agreeableness* T-score of 44.12 and a mean *Extraversion* T-score of 57.41, while the general population, by definition, has mean T-scores of 50 for both. This information is helpful when assessing individual pilots, as it places them in the proper context relative to their peers. The ALAPS may not be useful to those in the civilian sectors of aviation due to Federal law (the Americans with Disabilities Act) concerns, as it can be used to diagnose psychopathology in addition to measuring desirable personality traits. The problem would be administering it as part of a select-in procedure and violating Federal law by asking select-out type questions before extending a conditional employment offer. Further, it may be problematic as a selection tool due to the availability of the test manual (Retzlaff et al., 1997) in the open literature, encouraging coaching schemes, which could contribute to response inflation.

Potential Impact of New Technology on RPA Operator SAOC Requirements

RPA pilot SAOC requirements may be affected by mission objectives (e.g., manned-unmanned teaming, multi-RPA control), technology (e.g., automated take-off and landing, improved human-system interface design), and working conditions (e.g., work stressors such as shifts, number of hours, workload). It is likely that as technology advances, unmanned systems will become more autonomous, automated, and intelligent and more integrated with other manned and unmanned assets in a net-centric environment. Some tasks currently requiring manual control (take offs, landings, mission planning, sensor control) may be handled by automated systems, only requiring consent/approval by human operators. Decision aids (e.g., automatic target recognition, route planning, and timeline management) will enable the operator to assume more of a supervisory role in an integrated human-system team (van Breda, 2012). Technological developments may enable supervisory control of multiple RPAs or possibly swarms by a single operator. Under such conditions, mental and temporal workload will be high. SAOC requirements will focus on higher-order cognitive functioning. As aircraft autonomy increases, the need to manually control flight and psychomotor ability will decrease in importance. It is important that those responsible for human-system integration periodically examine the impact of changes in mission objectives and work environment and new technology on manpower, selection, and training requirements.

Discussion

Those responsible for human-system integration need to carefully consider the characteristics of human actors when developing or modifying systems. First, a job/task analysis must be done, including an analysis of cognitive, personality and other psychological characteristics needed for job success. Comparisons to the general population can be misleading. The use of specialized norms, when available and not prohibited by Section 106 of the Civil Rights Act of 1991, is highly recommended when assessing applicants as well as trained assets. Those responsible for human-system integration should also bear in mind the effects of changes in mission objectives and work environments and advances in technology on manpower, personnel, and training requirements. People, unlike machines, are prone to put their best foot forward (engage in response inflation) in an effort to influence an observer.

References

Carretta, T. R. (2005). *Development and validation of the Test of Basic Aviation Skills (TBAS),* AFRL-HE-WP-TR-2005-0172. Wright-Patterson AFB, OH, Air Force Research Laboratory, Human Effectiveness Directorate, Warfighter Interface Division.

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., & 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., Rose, M. R., & Bruskiewicz, K. T. (in press). Selection methods for remotely piloted aircraft systems operators. In N.J. Cooke, L. Rowe, & W. R. Bennett (Eds.). *Remotely piloted aircraft: A human systems integration perspective.* NY: Wiley.

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.

Chappelle, W., McDonald, K., Heaton, J. N., Thompson, W., & Haynes, J. (2012). Neuropsychological and personality attributes distinguishing high vs. low training performance of MQ-1B pilot trainees. *Aviation, Space, and Environmental Medicine, 83,* 261-262.

Costa, P. T., Jr., & McCrae, R. R. (1989). *The NEO Personality Inventory manual*. Odessa, FL: Psychological Assessment Resources.

Costa, P. T., Jr., & McCrae, R. R. (1992). "'Normal' personality inventories in clinical assessment: General requirements and the potential for using the NEO Personality Inventory": Reply. *Psychological Assessment*, *4*, 20-22.

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.

King, R. E., Barto, E., Ree, M. J., & Teachout, M. S. (2011). *Compilation of pilot personality norms* (Technical Report No. AFRL-SA-WP-TR-2011-0008). Wright-Patterson AFB, OH: U.S. Air Force School of Aerospace Medicine.

King, R. E., & Flynn, C.F. (1995). Defining and measuring the "Right Stuff": Neuropsychiatrically enhanced flight screening (N-EFS). *Aviation, Space, and Environmental Medicine, 66*, 951-956.

McHenry, J. J., Hough, L. M., Toquam, J. L., Hanson, M. A., & Ashworth, S. (1990). Project A validity results: The relationship between predictor and criterion domains. *Personnel Psychology, 43*, 335-354.

Parasuraman, R., & Byrne, E. A. (2003). Automation and human performance in aviation. 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. 311-356). Mahwah, NJ: Erlbaum.

Patrick, J. (2003). Trends and contexts of pilot training. 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. 398-434). Mahwah, NJ: Erlbaum.

Powell, D., Kaplan, E., Whitla, D., Weintraub, S., Catlin, R., & Funkenstein, H. (2004). *MicroCog: Assessment of cognitive functioning, Windows ed.* San Antonio, TX: Pearson.

Retzlaff, P. D., Callister, J. D., & King, R. E. (1997). *The Armstrong Laboratory Aviation Personality Survey (ALAPS): Norming and cross-validation* (Technical Report No. AL/AO-TR-1997-0099). Brooks AFB, TX: Armstrong Laboratory.

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,* 35-52.

Smallwood, T., & Fraser, M. (1995). *The airline training pilot.* Brookfield, VT: Ashgate.

United States Air Force (2011). *Medical examinations and standards* (AFI-48-123). Washington, DC: Author.

United States Air Force (2014). *Air Force human system integration handbook.* Brooks City Base, TX: 711 HPW/WPO. Retrieved December 18, 2014 from http://www.wpafb.af.mil/shared/media/document/AFD-090121-054.pdf.

van Breda (Ed.) ((2012). *Supervisory control of multiple unmanned systems – methodologies and enabling human-robot interface technologies, TR*-HFM-170. North Atlantic Treaty Organization Research and Technology Organization.

Williams, H. P., Carretta, T. R., Kirkendall, C.D., Barron, L. G., Stewart, J. E, & Rose, M. R. (2014). Selection for UAS Personnel (SUPer) phase I report: Identification of critical skills, abilities, and other characteristics and recommendations for test battery development, NAMRU-D Technical Report 15-16. Wright-Patterson AFB, OH: Naval Aeromedical Research Unit – Dayton.