2015

Panel on Cross-Cultural Pilot Selection

Diane L. Damos

Mark R. Rose

Monica Martinussen

Jan L. Lorenz

PANEL ON CROSS-CULTURAL PILOT SELECTION

Diane L. Damos
Damos Aviation Services, Inc.
Gurnee, IL
Mark R. Rose
Air Force Personnel Center
Randolph Air Force Base, TX
Monica Martinussen
The Norwegian Defense University College and
RKBU- Nord, Faculty of Health Sciences, UiT The Arctic University
Tromsø, Norway
Jan L.Lorenz
German Aerospace Center (DLR), Institute of Aerospace Medicine
Hamburg, Germany

Airlines in developing nations often face a pilot shortage. Airlines in such countries either must recruit experienced pilots worldwide or aid in the establishment of national flying schools. The first presentation explains some of the issues the airlines encounter in performing worldwide background screenings on experienced pilots. Selecting either experienced pilots or cadets from a multi-cultural, multi-lingual applicant pool is challenging. The second presentation discusses data comparing native versus immigrant cadet applicants on cognitive tests. The third presentation describes problems associated with adapting a personality assessment to a new culture and language and the resultant predictive validities. Selecting cadets for a national flying school in a multi-cultural developing nation presents additional challenges. The fourth presentation discusses some difficulties in assessing social skills in cadets in a multi-cultural developing country.

Civilian aviation is facing a shortage of both experienced pilots and young people who are interested in flying as a career. The shortage of experienced pilots is particularly acute in parts of Asia and the Middle East. Pilot shortages force airlines in the affected areas to recruit experienced pilots worldwide. Some airlines in the Middle East and Asia now have pilots from more than 50 countries who speak over 30 different native languages. Constructing a valid pilot selection system in a developed country is difficult when the applicants are ethnically mixed with a significant proportion speaking the official or dominant language non-natively. Constructing a selection system for an airline in a developing country is significantly more difficult when the applicant pool is multi-ethnic and multi-lingual.

Airlines in developing nations face several hurdles in developing both an appropriate screening process and a selection process with significant predictive validity. Criminal background checks and examinations of an applicant's driving record often are time-consuming and expensive and sometimes cannot be completed because of the privacy laws of the applicant's home country. The construction of a valid selection system may be hampered by the lack of appropriate selection instruments. Many, if not most, validated selection instruments are administered in a Western European language. Candidates who do not speak a Western European language natively may be at a disadvantage compared to native speakers. Personality tests are even more problematic because they are developed for a specific culture. Using these tests to assess candidates from other cultures may result in misleading results.

A long-term solution for many developing nations is to establish a national flight training school that can meet the country's need for trained pilots. Identifying young men and woman who are likely to be successful in a flying school is difficult for some of the reasons described above. An additional issue, however, is that some developing countries are themselves multi-cultural and multi-lingual.

The four sections comprising this paper describe some of the challenges of background screening and cognitive and personality assessments in multi-cultural environments. The first section deals with international background screening issues for experienced pilots. The next two sections discuss the challenges of using cognitive and personality assessment in multi-lingual, multi-cultural environments. The last section describes the development of a selection system for a cadet program in a multi-cultural nation with no locally validated selection instruments.

Multi-Cultural Pilot Screening Issues

Airlines typically begin the hiring process with background screening. The general purpose of screening is to ensure that an applicant has an acceptable employment history, the required experience minimums, no disqualifying legal events in his/her background, and the educational requirements. To assess the applicant's employment history, airlines typically ask for a list of prior employers and employment dates. The airline also may ask for letters of reference from the pilot's immediate supervisor or from other pilots who are familiar with the applicant. To verify the applicant's flight experience, airlines ask for copies of the pilot's licenses, certificates, and logbooks. Searches for legal issues may take a variety of forms. Airlines will search the pilot's initial civil license issuing authority's (CAA) accident and incident database. Similarly, if the pilot has been employed by different national airlines, the databases of each CAA will be searched. National databases of driving records also will be searched to determine if the applicant was ever convicted of Driving Under the Influence (DUI). If the carrier demands a specific level of education, the applicant may be asked to produce a copy of a diploma.

Screening of domestic applicants is usually straight forward for an airline in a developed country. The applicant's employment history can be relatively easily verified. Current employees who write letters of reference can be contacted. The CAA can be contacted to verify licenses and certificates. A national driving database often is readily available, as is the CAA's accident and incident database. Diplomas often can be verified with a telephone call. If the airline lacks sufficient in-house resources to perform the screening in a timely manner, it may hire specialized companies to perform background searches.

Nevertheless, airlines that hire domestically do encounter some problems. Of these, time may be the biggest issue. For example, the US is currently experiencing a pilot shortage at the regional airline level. Consequently, applicants may be hired and begin initial training before the results of the criminal background and DUI checks are available. Additionally, as the pilot shortage worsens, airlines increasingly must recruit internationally. Smaller airlines may lack the resources to vet international candidates. In the US, this has led the regional airlines to require a prior work history in the US and establish proof of residence. In some cases, the airline may only vet to the extent its resources allow.

Screening for airlines in developing nations is far more problematic. The government may or may not perform an extensive criminal background check before issuing a work permit. The airline may be unable to obtain all of the information from the applicant's CAA; some CAAs do not release pilot license and certification information and accident/incident data to foreign airlines. References from immediate supervisors may be particularly difficult to verify because of high pilot turnover rates at some new airlines in developing nations. An applicant's driving record is often ignored because of the expense of searching national databases, especially if the applicant has worked in several countries.

Multi-Cultural Cognitive Ability Testing

Cognitive ability testing has been a worldwide mainstay of pilot selection systems for more than half a century. Job analyses and validation studies generally support use of cognitive ability testing for evaluating pilot aptitude. However, few cognitive ability tests would meet international testing guidelines for interchangeable use across countries or cultures (International Test Commission, 2001). As noted by Ryan and Tippins (2009), validating a selection tool in one country is a difficult task; the task increases substantially when the process involves multiple countries and cultures. A partial list of factors that may confound the meaning and validity of cognitive test scores across countries includes cultural influences, constructs assessed, specific measures (e.g., item content, item difficulty), translation quality, scoring, study design, method of analysis, abilities of the sample, and the criteria.

This paper presents data from two groups of pilot applicants to the USAF who completed the Air Force Officer Qualifying Test (AFOQT) and the Test of Basic Aviation Skills (TBAS). The first group of consisted of applicants born in the US. The second group consisted of US citizens born outside the country or foreign nationals who later became naturalized US citizens. Because the US is a multi-ethnic, multi-cultural society, membership in these two groups is fuzzy, i.e., applicants have a degree of membership in each of the two groups.

Only the results from AFOQT and TBAS tests that assess constructs commonly included in pilot selection batteries—quantitative ability (Math Knowledge and Arithmetic Reasoning), verbal ability (Verbal Analogies and Word Knowledge), spatial ability (Rotated Blocks and Directional Orientation), and motivation (Aviation Information and Instrument Comprehension)—will be reported here. Table 1 presents descriptive data and correlations between scores on the test and pass/fail with Initial Flight Screening (IFS) and a comparison between the two groups. The results indicate that the cognitive measures (first six rows of Table 1) generally are valid predictors of IFS completion for applicants born inside and outside the US. Cohen's *d* values indicate that average between-group score differences are generally small (.12 to .31), and correlations corrected for dichotomization remain relatively larger for non-US born groups even when taking into account this group's higher IFS attrition rate (24.4% versus 9.3%).

Table 1.
Cognitive Ability Correlations with IFS Graduation/Elimination Across Cultural Groups.

	US (*n* = 4,288)				**Outside US (*n* = 164)**				
Measure	**Mean**	***SD***	**Obs *r***	**Dichot corr *r***	**Mean**	***SD***	**Obs *r***	**Dichot corr *r***	**Cohen's *d***
Math Knowledge	18.53	4.50	.06***	.10	17.80	4.93	.16*	.22	0.16
Arith Reasoning	18.68	4.41	.07***	.12	17.73	4.75	.14	.19	0.22
Verbal Analogies	18.07	3.41	.04*	.07	17.61	3.58	.15	.21	0.13
Word Knowledge	17.65	4.61	.03	.05	16.91	4.88	.16*	.22	0.16
Rotated Blocks	11.56	2.65	.10***	.17	10.74	3.15	.13	.18	0.31
Directional Orientation	0.31	0.88	.15***	.26	0.19	0.81	.25**	.34	0.14
Aviation Information	14.64	3.80	.25***	.44	14.18	4.09	.33***	.45	0.12
Instrument Comprehension	17.04	3.08	.21***	.37	16.45	3.82	.36***	.49	0.19

Note. *$p<.05$ **$p<.01$ ***$p<.001$; Obs *r* = observed point-biserial correlation coefficient; Dichot. corr *r* = observed correlation coefficient corrected for dichotomization of the criterion

Thus, these findings suggest that the same cognitive ability tests may be used to evaluate candidates with different country and cultural backgrounds. Measures of motivation, such as the Aviation Information Test, also appear to be useful predictors for some multi-cultural applicant populations.

Cross-Cultural Use of Trait-Based Personality Measures for Pilot Selection

Trait-based personality tests are frequently used as part of the selection system for both civil and military pilots. Some of these tests are developed for pilot selection, whereas others are measures developed for assessing personality traits in the normal population and are typically based on the Big-five model of personality. How the test results are used may vary between organizations and countries. Sometimes, the test results are used in addition to cognitive ability tests, whereas in other contexts the test results are used in combination with other types of information collected during an interview to assist in hiring decisions. International guidelines for tests and test use for personnel selection (e.g., European Federation of Psychologist's Associations (EFPA), http://www.efpa.eu/professional-development/assessment) state that selection tests should demonstrate reliability and predictive validity in addition to have appropriate norms and consideration for fairness.

Using a personality measure developed in one cultural context for assessing candidates from different language and cultural backgrounds raises a host of issues. Chief among these are problems with translation, appropriate norms, and the effect of biases, such as social desirability, on scores. Additionally, the predictive validity of such a test may decrease when it is used cross-culturally. Several studies have examined the equivalence of Big-five measures in different cultures and there is evidence supporting the five-factor structure across different language and cultural contexts (McCrae, 2002). However, there are also findings indicating that the reliability (in terms of internal consistency) and mean scores may differ between countries (McCrae, 2002). This raises cause for concern when testing applicants with different language and cultural backgrounds. Even for some Scandinavian

countries like Norway, Sweden, and Denmark---which are similar in many ways with a common history, culture and language—applicants may interpret sentences and adjectives in trait-based measures differently, sometimes resulting in invalid personality profiles. Another issue when using personality measures as part of the interview is that there may be cultural differences in self-presentation tactics e.g., in terms of asserting individual excellence and pointing out obstacles where some applicants may be viewed as underselling or overselling themselves compared to the perspective of the interviewer (Sandal et al., 2014). This highlights the importance of cultural competence among those conducting the assessments.

Personality tests in general have demonstrated a relatively modest predictive validity when used for pilot selection , whereas a more recent meta-analysis ($K = 8$) examining trait-based measures for military pilot selection found mean uncorrected correlations of -.15 and .13 for Neuroticism, and Extroversion, respectively (Campbell, Castaneda, & Pulos, 2009). These findings were supported in a study of US Air Force pilot trainees where some of the Big-five traits predicted training outcomes. However, the uncorrected correlations were generally small ($r < .11$) (Carretta, Teachout, Ree, Barto, King, & Michaels, 2014).

To conclude, there is some evidence supporting the predictive validity of trait-based measures, but these are mostly based on pilot samples from Europe or the US. In general the correlations are small, but the inclusion of an easy-to-administer, trait-based measure may result in incremental validity in the selection process. In addition, there are studies examining differences between pilot samples and the general population (Meško et al., 2013), and studies linking personality traits to team performance in aviation as well as to accident involvement (for an overview see Ganesh & Joseph, 2005). Taken together these findings indicate that pilot applicants differ from normative samples and that personality traits may indeed be important for pilot performance, even though documenting the predictive validity and generalizability of trait-based measures is still much needed.

Cross-Cultural Differences on Cognitive, Knowledge, and Assessment Center Measures Between Western European and Mauritius Cadet Applicants

A different approach that an airline in a developing country may consider when facing a pilot shortage is sponsoring its own cadet or ab initio program. This approach has a number of challenges. Some of these, however, are not that different from those encountered when hiring experienced pilots from varying nations: In developing states, for instance, even the first step of screening *within* its *own* population proves difficult at times. Frequently, educational and grading systems are not standardized and many young citizens seek educational opportunities abroad – either sponsored by their parents or governmental scholarships. The result is a confusing plethora of degrees, diplomas and grades, etc.

When the German Aerospace Center (DLR) was tasked with selecting viable candidates for the Mauritian Cadet Pilot Programme, we were faced with much the same problem that airlines hiring internationally face: Lacking an indigenous population, Mauritius is a truly multi-cultural nation. In addition to the Indo-Mauritian majority there are three prevalent ethnicities (Creole, Chinese, and European Mauritians) and a variety of practiced religions. How, then, can one make sure that all members of multi-cultural crews work towards creating a productive work environment by creating a shared ‘cockpit’-culture?

Consequently, in addition to employing internationally proven computerized knowledge and cognitive tests (e.g. Zierke, 2014; Maschke, Oubaid & Pecena, 2011) as well an aviation-specific personality test and a final interview, we decided to include an observed team task for the first time in over four decades of global assessment efforts. We would not have felt comfortable recommending applicants without having observed their behavior while working cooperatively in diverse teams.

We were, of course, aware of the possible calamities that might result when a translated task, originally developed and tried in a western European country, was used in a different cultural context. Culture cannot be experimentally varied and the construct culture itself is neither mono-causal nor does it have trivially discernible consequences. Yet, as others have argued, all tests that we used had been explicitly designed to measure inter-individual differences. As long as extraneous conditions are controlled and the methods are sufficiently sensitive and selective, employing such measures in varying (cross-)cultural settings appears warranted (Simon, 2006).

Therefore, we were confident that the task (distributing turns of duty with varying popularity among a group of First Officers) would consistently produce a wide range of relevant behavior and that our rating system

would be suited to measure this behavior adequately. Four dimensions of performance were assessed by independent raters that are significant for safety in aviation: Leadership, Cooperation, Communication and Stress Resistance.

Because of the heterogeneous ethnic composition of the Mauritian population, and the constant need to bridge these differences by means of cooperation and negotiation, and the fact that the original countries of Mauritian ethnicities mostly score high on Hofstede's (2011; Hofstede, Hofstede, & Minkov, 2010) collectivism dimension, we expected a more pronounced cooperative effort from Mauritians when compared to German applicants (Paul, Samarah, Seetharaman & Mykytyn Jr, 2004; McLeod, Lobel & Cox, 1996). In addition we expected some Mauritian applicants to display a tendency to behave more timidly when compared to Germans, resulting in a low average score on Communication.

To test these hypotheses, the results of 52 Mauritians (age M=24.73, SD=2.20, 92.3% male) were compared with those of an analogous group of ab initio candidates from Germany (n=860, age M=21.16, SD=2.04; 90.5% male). The results are shown in Table 2. In contrast to our expectation, the Mauritian applicants were just as communicative as the Germans, even though there was a slightly higher variance among the applicants. Yet, the major hypothesis – that Mauritians behave more cooperatively than Germans – was supported by our data.

Table 2.
Mean scores and standard deviations of Mauritian and German applicants in Cooperation and Communication

	Germans *n*=860		Mauritians *n*=52	
	M	SD	M	SD
Cooperation	3.32	.72	3.77	.91
Communication	3.01	.87	3.06	1.18

Notes. Mean differences: Cooperation: t=3.53 p<.001, g=.62; Communication: t=0.28 p=.783, g=.06

For practitioners, it is important to note that our team task was very well suited for use in a diverse cultural setting. Both the task itself and the range of its rating system provided markedly selective results and , thus, a good basis for decisions and recommendations. In this specific example, we simply had to shift our focus *within* the existing methodology – from a given cooperation towards an emphasis on leadership skills.

Judging from this experience, we would encourage the application of work-related team tasks to directly observe the behavior in diverse teams as a measure for intercultural cooperation and suitability for a multi-cultural cockpit environment. Of course, every effort must be made to create tasks which are culturally fair and unequivocal for all participants. In our opinion, the inclusion of a team task extends the validity of the whole assessment process, e.g., by contributing hypotheses for subsequent interviews.

Conclusion

Because of pilot shortages in many parts of the world, airlines are forced either to recruit experienced pilots internationally or to develop their own cadet programs. Screening foreign experienced pilots is difficult because of privacy issues of some countries and the cost of searching numerous databases. Selecting pilots from multiple countries is problematic because it requires valid selection instruments that can be used cross-culturally. Currently, cognitive, personality, and motivational-assessment instruments show some cross-cultural validity but need further development. Team performance assessments may provide a valuable tool for determining how well an individual can function in a multi-cultural environment.

References

Campbell, J. S., Castaneda, M., & Pulos, S. (2009). Meta-analysis of personality assessments as predictors of military aviation training success. *The International Journal of Aviation Psychology, 20*, 92-109. doi:0.1080/10508410903415872

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. *The International Journal of Aviation Psychology, 24*, 247-264. doi: 10.1080/10508414.2014.949200

Ganesh, A., & Joseph, *C.* (2005). Personality studies in aircrew: An overview. *Indian Journal of Aerospace Medicine, 49*, 54-62.

Hofstede, G. (2011). Dimensionalizing Cultures: The Hofstede Model in Context. *Online Readings in Psychology and Culture, 2*(1). Retrieved from http://dx.doi.org/10.9707/2307-0919.1014

Hofstede, G., Hofstede, G.J., & Minkov, M. (2010). *Cultures and Organizations: Software of the Mind*, (3rd ed.), NY: McGraw Hill.

International Test Commission (2001). *International Test Commission guidelines for test adaptation*. London: Author.

Maschke, P., Oubaid, V., & Pecena, Y. (2011). How do astronaut candidate profiles differ from airline pilot profiles? Results from the 2008/2009 ESA astronaut selection. *Aviation Psychology and Applied Human Factors, 1*, 38-44.

McCrae, R. R. (2002). Cross-cultural research on the five-factor model of personality. *Online Readings in Psychology and Culture, 4*(4). Retrieved from http://dx.doi.org/10.9707/2307-0919.1038

McLeod, P. L., Lobel, S. A., & Cox, T. H. (1996). Ethnic diversity and creativity in small groups. *Small Group Research, 27*, 248-264.

Meško, M., Karpljuk, D., Štok, Z. M.,Videmšek, M., Bertoncel, T., Bertoncelj, A., & Podbregar, I. (2013). Motor abilities and psychological characteristics of Slovene military pilots. *The International Journal of Aviation Psychology, 23*, 306-318. doi: 10.1080/10508414.2013.833750

Paul, S., Samarah, I. M., Seetharaman, P., & Mykytyn Jr, P. P. (2004). An empirical investigation of collaborative conflict management style in group support system-based global virtual teams. *Journal of Management Information Systems, 21*, 185-222.

Ryan, A.M. & Tippins, N. (2009). Designing and implementing global selection systems. Malden, MA: Wiley.

Sandal, G. M., van de Vijver, F., Bye, H. H., Sam, D.L., Amponsah B., Cakar, N.,,..Tien-Lun Sun, C. (2014). Intended self-presentation tactics in job interviews: A 10-country study. *Journal of Cross-Cultural Psychology*, 45, 939-958. doi: 10.1177/0022022114532353

Simon, P. (2006). The solution of fundamental methodological problems in cross-cultural psychology by guaranteeing the equivalence of measurements. In J. Straub, D. Weidemann, C. Kölbl & B. Zielke (Eds.), *Pursuit of meaning: Advances in cultural and cross-cultural psychology* (pp. 269-292). Bielefeld: Transcript.

Zierke, O. (2014). Predictive validity of knowledge tests for pilot training outcome. *Aviation Psychology and Applied Human Factors*, *4*, 98-105. doi: 10.1027/2192-0923/a000061