PILOT ABILITIES AND PERFORMANCE

Deborah A. Boehm-Davis, Robert W. Holt, and Jeffrey T. Hansberger
George Mason University
Fairfax, Virginia

ABSTRACT

Commercial pilots completed a series of tests measuring individual differences in cognitive, perceptual, and verbal skill. These data were used to describe pilot abilities relative to samples upon which the tests were normed. Further, the data were used to examine the relationships between these measured abilities and performance on maneuver validation and line-oriented evaluations collected during annual recurrent training. The data suggest that (1) commercial pilots have higher levels of reasoning and perceptual abilities than the comparison groups of naval recruits or young males and slightly higher levels of verbal ability; and (2) these abilities are related to maneuver validation and line-oriented evaluation (LOE) performance.

INTRODUCTION

In the aviation industry, pilot performance is imperative because of the possible dire consequences when pilots do not perform to preset standards. In order to prevent accidents and improve pilot training and selection, it is necessary to understand what factors are involved in pilot performance. Each pilot on the line enters an aircraft with varying levels of the same basic abilities. Through the use of many of these abilities, the complex task of safely and efficiently flying the aircraft is attempted by the pilot. By examining the relationship of a pilot's abilities to their flying performance, an understanding of exactly what abilities are needed for optimal performance in a cockpit may be found. These abilities can then be targeted for specific training and/or selection among the pilot population.

Prior research in the aviation domain has primarily been concentrated in a military setting with fewer studies being involved with the commercial aviation population. It is important to study both since they represent different flying environments as well as different approaches to training and selection. This study addresses this issue by examining commercial pilots at a regional airline. Another important aspect that has been rare in past research is obtaining overall flight performance ratings. This is in contrast to much of the research that has used varying pieces of pilot performance like situational awareness (Endsley, 1994) or workload management (Raby & Wickens, 1994) that captures only part of the pilot's flight task. It is also important to obtain performance data in the most realistic scenario and environment as possible. The use of actual training and line performance data would provide the most valid and generalizable results for research. Data from the participating regional's line-oriented evaluation (LOE) and maneuver validation were used as measures of pilot performance in this research.

In order to explore the relationship between individual differences among pilots and their performance, an understanding of the individual differences in the pilot population is needed. Few studies have compared the perceptual abilities of pilots to the general population and fewer have compared their logical and verbal abilities to the general population. Research that has been conducted on spatial abilities (Dror, Kosslyn, & Waag, 1993; Temme, Still, & Fatcheric, 1991) has shown that pilots possesses better spatial abilities for some tasks than the general population. However, there is less known about other abilities such as logical and verbal abilities that may have possible crew resource management (CRM) implications among the pilot population. It is also important to examine the relationship between the abilities to assess whether they are actually measuring separate abilities and their relationship to one another. By understanding this relationship, training departments can design their programs to address the key and integral abilities involved.

Once the individual differences of the pilots have been assessed, the underlying question can be addressed, what is the relationship between individual differences of pilots and their performance in actual training scenarios? There is little research in the commercial aviation domain examining the pilot's overall flight performance in training events like the LOE and maneuver validation and actual line performance data.

The implications of thoroughly understanding what individual differences among pilots lead to superior performance in the cockpit could have a great impact on pilot selection and training if these abilities can be trained. There is evidence that *particular* perceptual abilities can be improved through training (Lohman & Nichols, 1991; Baenninger & Newcombe, 1989; Embertson, 1987; Yates, 1986). There has also been research that suggests that verbal (Sahu & Swain, 1985) and reasoning abilities (Schaie, Willis, Hertzog, & Schulenberg, 1987) can be increased by training as well.

The objective of this study is to explore the relationship between individual differences in reasoning, perceptual, and verbal abilities to the pilot's actual performance in high fidelity flight training situations.

METHOD

Participants

The participants were 258 pilots employed by a regional airline. Based on demographic data questionnaire completed by a subset of 62 of the pilots, the mean age of the pilots was 34. The pilots had been employed with this airline for a mean of 4.3 years. All of the pilots completing the demographic questionnaire were male (although there was a small number of females in the larger sample). All of the pilots in the larger sample were assigned to one of two fleets within the carrier's operation.

Materials

The set of materials consisted of a background questionnaire (of approximately 250 items) and a set of tests designed to measure a broad range of pilot abilities. The tests were all drawn from the ETS set of factor-referenced cognitive tests (Ekstrom, French, Harman, and Derman, 1976). Specifically, tests were selected to measure: perceptual recognition, selective perception, logical reasoning, word fluency, and verbal comprehension/vocabulary. Perceptual recognition reflects an individual's ability to accurately recognize ambiguous stimuli. The Snowy Pictures test, which measures "speed of closure", or the ability to quickly unite different pieces of a perceptual field into a single perceived object, was used to measure this factor.

Selective perception reflects an individual's ability to search for a particular visual pattern in a distracting perceptual field. The Hidden Patterns test, which measures "flexibility of closure", or the ability to focus on one part of a visual percept and recognize it among distracting material in the visual field, was selected to measure this factor.

Logical reasoning reflects the ability to evaluate the logical correctness of possible conclusions for a set of information. The Inference test, which measures the ability to reason from premises to conclusions, was used to measure this factor.

Word fluency refers to the general ability to use phonetic cues to produce appropriate words. The Word Beginnings and Word Endings tests were used to measure this factor. The Word Beginnings test presents beginning word stems, and participants needed to produce as many words from these beginnings as possible. The Word Endings test was identical, except that it used end fragments as stimuli.

Verbal comprehension reflects an individual's ability to understand language. Vocabulary ability, which is part of verbal comprehension, was used as the measure of comprehension. The Extended Range Vocabulary test was selected to measure this ability. This test provides participants with a target word and asks them to select the word that most closely matches the target word in meaning from a set of five alternatives.

Procedure

The individual difference tests were administered either at the end of CRM training (for one fleet) or after recurrent training (for another). Pilots who completed the individual difference tests were given the demographic questionnaire and asked to return it to the training center. Pilots received several reminders to return the questionnaire; however, only a small number of completed questionnaires were returned. Sometime during the one-year period following the collection of the paper and pencil materials, pilots

were observed in one maneuver validation and one (LOE) simulation session that were required for recurrent training. During the initial simulator session, the pilot's ability to maneuver the aircraft under varying conditions was observed and an overall rating was assigned to each pilot on the basis of their performance. The LOE consisted of a number of “legs”, which were further broken up into event sets. Event sets are organized around a “trigger” event, which provides an opportunity for the pilots to respond and for the evaluators to assess the capabilities of the pilots as they respond to the event. Within each event set, the instructor rated the overall performance (based on their technical and crew resource management (CRM) performance) for both the Captain and First Officer, noting which pilot was flying the aircraft.

RESULTS

Differences between Pilots and the General Population

The individual difference test scores were calculated for each pilot and plotted in histogram form for each individual test. Comparison group data (means and standard deviations) were selected for each test from the manual describing the tests (Ekstrom, French, Harman and Derman, 1976); the comparison groups chosen were selected to provide the best match to our sample and consisted of either servicemen (Army or Navy), or college students.

Perceptual Recognition. Figure 1 shows the distribution of test scores on the Snowy Pictures test for the sample of 238 pilots. This figure is representative of the set of graphs constructed for each of the abilities. The figure also shows the mean score (5.7) earned by the comparison sample of naval recruits as given in Ekstrom *et al* (1976). A *t*-test indicates that the mean of the pilot group (13.1) is significantly higher than that of the comparison group ($t(775) = 30.1$, $p < .01$).

Selective Perception. The pilot’s average score on the Hidden Patterns test (mean = 205.8) is significantly higher than that of the college males comparison group (mean = 148) suggesting a superior ability to locate specific visual patterns in the midst of a complex visual field ($t(490) = 16.06$, $p < .01$).

Logical Reasoning. The average score for the pilots on the Inference test (mean - 14.3) was significantly higher ($t(490) = 7.03$, $p < .01$) than that of the college male comparison group (mean = 12), suggesting a superior ability to reason from premises to conclusions.

Word Fluency. The pilots in our sample scored slightly higher on average in their ability to produce words based on beginning and ending word stems (56.8) than the average (52.5) of college student and Army enlistee comparison samples. Although this difference in average scores is reliable ($t(410) = 3.56$, $p < .01$), it is small (on average, less than 4 words).

Pilots’ Vocabulary. As with word fluency, there was a reliable ($t(490) = 2.44$, $p < .01$), but small difference between the average score for Army enlistees (mean = 19.7) and pilots (mean = 21.35). On average, pilots were able to match a definition to the appropriate word two more times than the comparison group.

Relationships among Abilities

Using the sample of 238 pilots for whom we had individual differences data, we examined the co-variation among the set of abilities that we had measured. Figure 2 shows the joint variability (R^2) among each the five measured abilities. The pattern of overlapping variance suggests a “chain” of related abilities. However, each ability test also measures unique variance, so each also appears to contribute independently to our ability to measure performance.

Relationship between Ability Measures and Performance

Data are available for a sample of pilots who participated in a recurrent training session that included maneuver validation (data available from 121 pilots) and an LOE (data available from 105 pilots) within a year of completing the individual difference tests. During the maneuver validation, overall performance scores were given. In the LOE, an overall performance score was given for each segment (or leg) of the

LOE. In each case, the data were identified by a personal identification number (PIN), as well as by whether the individual was the captain or first officer and whether the pilot was flying (PF) or not flying (PNF).

The data from the abilities tests were correlated with maneuver validation and LOE scores. None of the abilities measures was significantly correlated with overall maneuver validation proficiency when the data for Captains and First Officers (when they were the pilot flying) were analyzed separately. When the data for the Captains and First Officers were combined, perceptual recognition, as measured by the Hidden Patterns test, came close to reaching an acceptable level of significance ($r(121) = .18$, $p = .052$). This suggests that the only component of the abilities measures that potentially relates to aircraft control skills is the ability to recognize ambiguous stimuli.

For LOE performance, the data from each individual leg and event set were correlated with the abilities data. Across the 11 individual event sets, logical reasoning ability was significantly correlated with performance on five of the event sets (mean $r(105) = .20$, $p < .05$); for these legs, increased logical reasoning ability was associated with improved LOE performance. In this each of these event sets, there was an element of problem solving associated with dealing with an abnormal situation (e.g., determining that an abnormal reading on a guage is only a minor problem (leg 1, event set 3) or planning which airport to divert to due to an abnormal (leg 1, event set 4 and leg 2, event set 4). Perceptual recognition ability was associated with performance on four of the event sets (mean $r(105) = .22$, $p < .05$); these event sets contained an element of performance that required perceptual skills (e.g., checking guages during the pre-flight and dealing with a hot start (leg 1, event set 1) or recognizing that a diversion to another airport was necessary due to low fuel and traffic at the planned airport (leg 2, event set 4). Finally, selective perception, as measured by the Hidden Patterns test, was related to performance in three of the event sets (mean $r(105) = .18$, $p < .05$); these event sets contained situations requiring attention to perceptual information that may have been hidden in a complex background (e.g., following the guages associated with a hot start while preparing for departure (leg 1, event set 1) or watching out for traffic while taxiing to take-off (leg 2, event set 2).

In addition, the average performance score, across the event sets, was correlated with performance on the ability measures. In this analysis, only the Snowy Pictures test, which measures perceptual recognition, was found to be significantly correlated with LOE performance ($r(105) = .23$, $p < .05$).

DISCUSSION

This study was designed to describe the abilities of commercial pilots relative to the general population and to determine the relationship between these abilities and performance in the cockpit. An examination of five specific abilities measuring reasoning, perceptual, and verbal fluency factors found that commercial pilots are superior to comparison groups of servicemen and college students who were used to validate the tests. Although the pilots were superior on all abilities measured, their advantage was greatest for perceptual factors. This is interesting for a number of reasons. First, pilots for regional airlines, unlike pilots who fly for the services or the major carriers, were not selected by the airline on the basis of abilities testing. This suggests that there may be a process of self-selection in becoming a pilot. An alternative interpretation is that the daily exposure to perceptual tasks improves individual abilities. This is contrary to the view that these tests measure innate traits rather than trained abilities; however, this explanation makes it possible to consider the development of training programs for individuals low in this ability. It is also consistent with evidence that women, who generally score lower on spatial ability tests, benefit more from spatial training than men (Lohman & Nichols, 1991; Blatter, 1983).

A second goal of the study was to determine if there was a relationship between abilities and performance. Here, the data suggested some interesting relationships. Somewhat surprisingly, the measured abilities did not seem to be predictive of maneuver validation performance. Only one of the perceptual abilities tests (Snowy Pictures) was correlated (weakly) with performance in controlling the aircraft. This could be the result either of self-selection or of training. The comparisons of our pilots with "average" groups of people suggested that our pilots were higher in the measured abilities; this restriction of range may have

influenced the lack of correlation. Another possible explanation, though, is that the extensive number of hours flying required to become a commercial pilot reduces the effects of native abilities on performance.

This second explanation may be more likely as relationships were found between LOE performance and perceptual recognition, selective perception, and reasoning abilities. In each of these cases, increased perceptual and reasoning abilities were associated with better performance in event sets that contained abnormal conditions that required reading/checking guages in order to diagnose or solve problems. It was somewhat surprising that verbal fluency and vocabulary abilities were not related to performance in any of the individual event sets. Because pilots are required to communicate with each other through briefings, and to communicate with dispatch and air traffic control (ATC), we had anticipated that verbal abilities would be related to performance. It is possible that either our measures of verbal skills are not adequate, or verbal communications do not contribute strongly to the overall ratings of performance in LOEs.

Overall, this research suggests that pilots have superior perceptual and reasoning skills compared to the general population and that some of these skills are related to performance in a realistic environment. To the extent that these abilities can be trained, these data further suggest a starting point for developing training programs.

REFERENCES

Baenninger, M. and Newcombe, N. (1989). The role of experience in spatial test performance: A meta-analysis. *Sex-Roles,20*(5-6),327-344.

Blatter, P. (1983). Training in spatial ability: A test of Sherman's hypothesis. *Perceptual and Motor Skills, 57*(3 Pt 1), 987-992.

Dror, I. E., K., S. M., & Waag, W. L. (1993). Visual-spatial abilities of pilots. *Journal of Applied Psychology, 78*(5), 763-773.

Ekstrom, R. B., French, J. W., Harman, H. H., and Derman, D. (1976). *Manual for kit of factor-referenced cognitive tests*. Princeton, NJ: Educational Testing Service.

Embretson, S.E. (1986). Improving the measurement of spatial aptitude by dynamic testing. *Intelligence, 11*(4), 333-358.

Endsley, E.R. and Bolstad, C.A. (1994). Individual differences in pilot situation awareness. *The International Journal of Aviation Psychology, 4*(3), 241-264.

Lohman, D.F. and Nichols, P.D. (1990). Training spatial abilities: Effects of practice on rotation and synthesis tasks. *Learning and Individual Differences, 2*(1), 67-93.

Yates, L.G. (1986). Effect of visualization training on spatial ability test scores. *Journal of Mental Imagery, 10*(1), 81-91.

Schaie, W.K., Willis, S.L., Hertzog, C., & Schulenberg, J.E. (1987). Effects of cognitive training on primary mental ability structure. *Psychology and Aging, 2*(3), 233-242.

Sahu, S. and Swain, J.B. (1985). Effect of language training programme on low language achievers. *Perspectives in Psychological Researches 8*(2), 1-8.

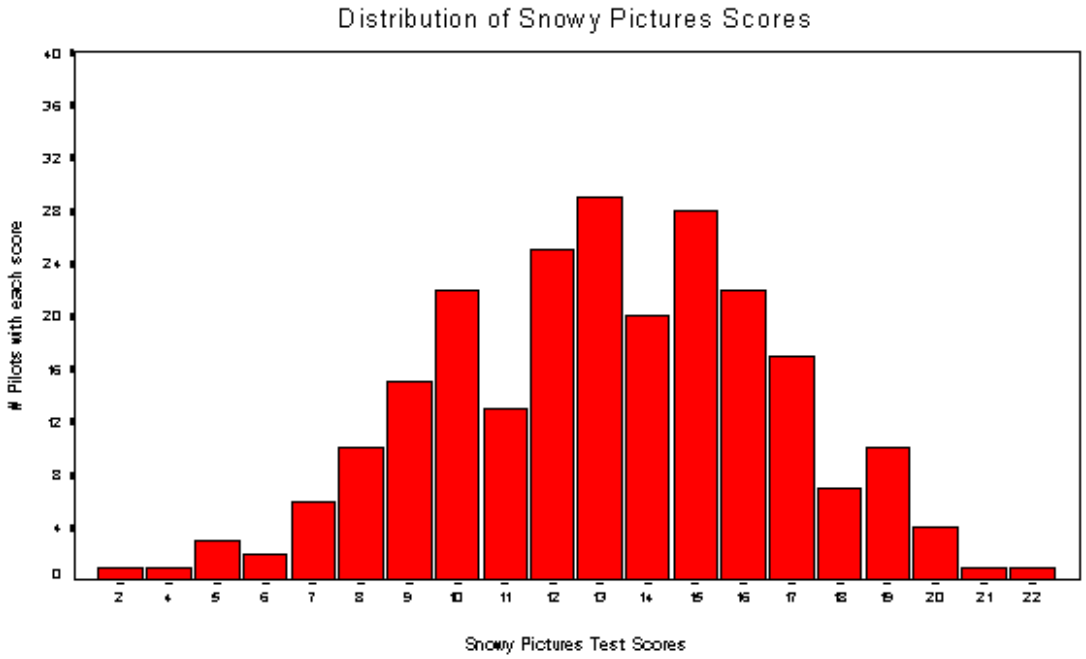

Figure 1. Histogram showing the distribution of pilots' scores on the Snowy Pictures test. The mean score from the distribution of scores for the naval recruits used to validate the test (Ekstrom *et al*, 1976) was 5.7.

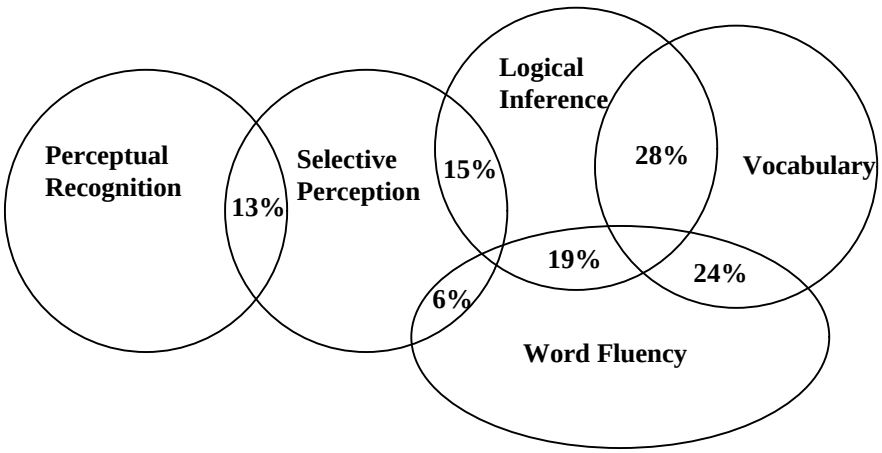

Figure 2. Variability (R^2) shared by each of the individual differences measures.