

Using The Table Reading Test As An Indicator For Success in Pilot Training

Anthony Mekhail, Mary Niemczyk, Jon W. Ulrich, and Merrill Karp
Arizona State University

ABSTRACT

The Table Reading Test (TRT) assesses an individual's perceptual speed, or how rapidly one absorbs and processes visual search information. The focus of this study was to determine the validity of the TRT scores as a predictor of pilot training success. A total of 116 subjects enrolled in an aviation program at a major university in the southwest were tested between the Fall 2005 and Fall 2008 semesters, inclusive. Their TRT scores were tested for correlations to one of several flight/academic performance criterion. The results of the analyses found the TRT best predicts: Time to Solo ($r = -0.228, p < 0.024$), Time to Private ($r = -0.754, p < 0.001$), and GPA ($r = 0.283, p < 0.002$).

INTRODUCTION

Many young adults pursue flight training with aspirations for careers in the military, commercial airlines, and/or with private entities. Some do so because of an interest in aviation, while others may do so because of extrinsic reasons, such as pressure from parents who may themselves be pilots. Because the pilot profession is a well respected and prestigious career path in many societies, some individuals set out to become pilots without fully realizing the many challenges that lay ahead of them.

One of the biggest challenges in becoming a pilot is the ever increasing cost of flight training. Flight training is quite expensive, potentially costing students tens of thousands of dollars over the course of their instruction. A major component of this expense is related to the fleet of aircraft that flight schools procure and maintain for training. In addition, flight schools must hire well-qualified flight instructors and may also provide ancillary facilities and services, such as hangars and computer simulators.

Not all students who begin flight training actually complete it. In the early stages of flight instruction, students may decide to discontinue training for a variety of reasons - some may lose interest, while others may determine that they cannot afford further training. In other cases, students may continue their training only to discover that they do not possess the aptitude necessary to become a pilot. This could result in frustration and disappointment along with a significant amount of debt.

To resolve this situation, it would be helpful to identify early on those individuals who possess the greatest potential to succeed in flight training. In order to properly identify pilot candidates with the greatest chances for success, individuals should be assessed before beginning training. The military, for example, uses paper-based and apparatus-type tests focusing on cognitive and psychomotor abilities. In contrast, commercial airlines hire pilots with different aviation backgrounds and experiences and, instead, test for personality and job sample measures (Goeters, 2004). While some airlines hire pilots from the military, pilots are also hired away from other carriers, while still others come directly from training schools (Carretta & Ree, 2000).

Although there are several different test instruments used for pilot selection, almost all measure cognitive ability, conscientiousness, or job knowledge. The focus of this study was to determine the validity of a paper-based assessment measure, the Table Reading Test (TRT) (Damos, 2004), to assess an individual's perceptual speed, or how rapidly one perceives and processes, in this case, visual information. Damos has previously validated the TRT for airline pilots; therefore the focus of this study was the validation of the test for *ab initio* pilots.

Statement of Purpose

For many years, the aviation industry has sought to determine the characteristics that constitute a successful pilot. This study focuses on determining the possible correlation of subject TRT scores to flight and academic performance measures, including the number of hours accumulated before solo, the number of hours accumulated prior to the FAA Private Pilot check ride, and an individual's Grade Point Average (GPA).

LITERATURE REVIEW

Pilot selection systems are designed to relate test scores to skills that are required to fly an aircraft. While selection assessment measures have been utilized for many years (Thorndike, 1949) and have included various constructs, the most essential task remains the same - to obtain suitable criterion measures to validate against performance.

Validity

Validity is the degree to which an instrument or test measures what it intends to measure (Carmines & Zeller, 1979) and is an overall evaluation of the extent by which observations and theories support interpretations derived from tests and instrument scores (Messick, 1995). However, validity is not a property of the test itself, but rather the interpretation of the test results. It is important to identify the intended purpose of the instrument, the intended subjects on which the instrument is to be used, and the specific conditions applicable to the instrument with regards to how and what inferences can be made from the test's results (Di Iorio, 2005). If validity cannot be shown, a study has very little value regardless of the relationships between test scores and performance (Carretta & Ree, 2000).

There are three types of validity: Content, Construct, and Criterion. Content validity tests whether an experimental or observational measurement reflects a particular area of content (Carmines & Zeller, 1979). That is, content validity tests whether the *content* of a test actually assesses all of the measures the test is intended to assess. Construct validity, on the other hand, pertains to determining exactly what a specific test actually measures. It allows one to *construct* and then test a hypothesis. Construct validity is used when no criterion or content is sufficiently accepted to define a subject that is being measured. Construct validity considers the relationship between an observed measure to its supporting theories (Carmines & Zeller, 1979). Finally, criterion validity looks at test scores and predicts future subject population performances. If the correlation between scores and performance on the criterion variable is insignificant, then the test is not useful. The purpose of criterion validity is to show that scores on the criterion variable can be predicted from test scores or the predictor variable.

While all three validity tests are important, *criterion validity* is particularly important in pilot selection activities. That is, criterion validity assesses the relationship between TRT scores and the two pilot flight performance measures as well as student GPAs. In this research, *pilot performance* was defined to be flight training hours to solo and flight training hours to obtain a private pilot's certificate.

Selection Systems

Any occupational selection process involves the screening of applicants, followed by choosing the applicants based on certain methods and criteria. The purpose of selection systems are to choose those

applicants that appear to have the most promise and potential to successfully perform job requirements, which allows training resources to be used more efficiently. Ideally, selection systems focus on job-relevant characteristics, when there is a discrepancy among the applicant's population, and when these characteristics cannot be trained (Goeters, 1998).

Previous Pilot Selection Studies

Pilot selection methods typically focus on cognitive and psychomotor ability. They are also designed to predict pass/fail of flight training but rarely are designed to predict operational performance (Damos, 1996). However, with the current trend of increasing crew-to-aircraft and crew-to-crew interfaces, other skills and abilities need to be investigated. Today, pilots must be able to monitor and cross-check flight parameters and aircraft systems, integrate information coming in from various channels, maintain situational awareness, communicate with crew members, and coordinate appropriate actions (Goeters, 1998). Modern selection test batteries must account not only for cognitive and motor skills but also personality and operational performance.

In the search for the ideal pilot selection system, a number of personnel measurement methods have been studied. Validation studies are required to establish the value and usefulness of tests designed to predict job performance. The current research seeks to validate the TRT; therefore, it is useful to review other pilot selection validation studies to determine what has already been completed. In this review of previous pilot selection studies, unless specified otherwise in the narrative, all reported r -values were found to be significant to the $p < 0.05$ level.

Military cognitive/psychomotor ability studies.

Henmon (1919) evaluated the Thorndike Intelligence Test in the selection of Army Air Service Pilots. The test was administered to 150 pilots – 50 who were rated as very good, 50 who were rated as very poor, and 50 pilots of unknown flying ability. Henmon found a correlation of $r = 0.35$ between the Thorndike Mental Alertness Test scores and instructors' ratings of flying ability.

Fiske (1947) conducted an evaluation of a mechanical comprehension test as well as the Wonderlic's Personnel Test (WPT), which measures general intelligence, used in the selection of U.S. Navy pilots. In three samples of approximately 2,000 subjects each, Fiske (1947) reported that the mechanical comprehension test revealed higher correlations than the WPT with regards to pass/fail criterion. All samples showed similar results. Mechanical comprehension characteristics had higher correlations than general intelligence. Similarly, studies by the U.S. Army Air Corps found results commensurate with the mechanical comprehension test to those obtained in the Navy study (Hunter & Burke, 1995).

Signori (1949) studied selection tests used by the Royal Canadian Air Force during World War II. Five general cognitive tests were administered and were correlated to pass/fail results in training: mechanical reasoning test ($r = 0.21$), practical mechanical ability ($r = 0.17$), and general mental ability ($r = 0.06$). In a study of U.S. Navy pilot selection, Bair, Lockman, and Martoccia (1956) reported correlations between intellect, spatial ability, and clerical functions and a training performance rating ranging from $r = 0.10$ to 0.26 .

Want (1962) conducted a study of Royal Australian Air Force pilot candidates. One of the tests measured instrument comprehension, requiring candidates to understand and interpret aircraft instruments. This instrument comprehension test had the highest correlation with training criterion among all the tests he administered ($r = 0.46$). The other tests were dial reading, silhouette recognition, general information, mathematics, and general science, each finding similar correlations ($r \approx 0.25$).

North and Griffin (1977) reviewed naval aviator selection from 1917 to 1977 that included 2,109 pilot candidates. Individual components of the Navy's selection battery were compared to pass/fail criterion. The mechanical comprehension subtest and biographical inventory had correlations of $r = 0.19$ with pass/fail criterion, while the spatial apperceptions subtest had a correlation of $r = 0.11$ with biographical inventory.

An important component in United States Air Force (USAF) pilot selection is the Air Force Officer Qualifying Test (AFOQT). This is used to select civilian or prior service applicants and to classify commissioned officer applicants into aircrew specialties (Carretta & Ree, 1995). The AFOQT is a battery of 16 tests that assesses five ability domains: Verbal, quantitative, spatial, aircrew interest/aptitude, and perceptual speed. USAF pilot trainees must complete a 53-week flight training job knowledge and job skills program. A total of 7,563 USAF officers were part of a study conducted by Carretta and Ree (1995) to validate the prediction of the pilot training criteria based on AFOQT scores. These subjects attended undergraduate pilot training and all had been tested on the AFOQT.

Pilot training is broken down into three phases. Phase 1 consists of classroom training (ground school) in aeronautics. Phase 2 consists of initial training in a subsonic jet aircraft, while Phase 3 consists of advance training in a transonic jet aircraft. The criteria used in this study was aeronautics average (Phase 1), daily flying and check flight averages (Phase 2), and daily flying and check flight averages (Phase 3). Arithmetic reasoning (quantitative) was the best predictor for Phase 1 ($r = 0.20$). Aviation information (aircrew interest/aptitude) was the best predictor for Phase 2 daily flight averages ($r = 0.25$), while instrument comprehension (aircrew interest/aptitude) was the best predictor of Phase 2 check flight averages ($r = 0.19$). Scale reading (perceptual speed) was the best predictor of Phase 3 daily flight averages ($r = 0.08$). Scale reading was again a best predictor of the Phase 3 check flight averages ($r = 0.13$).

Another important component in USAF pilot selection is the Basic Attributes Test (BAT). The BAT is a computerized battery of psychomotor, cognitive (speed of information processing), and personality (attitude toward risk) tests (Carretta & Ree, 1996). The BAT contributes to a pilot selection composite used by the USAF called the Pilot Candidate Selection Method (PCSM) (Carretta & Ree, 1998). PCSM scores have been shown to predict several measures of pilot training performance. Some of these include passing/failing pilot training and class rank with BAT psychomotor test scores having correlations of $r = 0.15$ and $r = 0.16$, respectively (Griffin & Koonce, 1996).

A study conducted by Bailey and Woodhead (1996) assessed the Royal Air Force (RAF) aircrew selection methods. The RAF pilot aptitude test is a computer-based test system that considers five domains: Control velocity, sensory motor apparatus, instrument comprehension, vigilance, and digit recall. Based on the results of their study, Bailey and Woodhead reported that the predictive validity of the pilot aptitude composite scores against basic flight training had a correlation coefficient of $r = 0.52$. The RAF system is commercially available and has been purchased by several civilian airlines and military services. This system has also been used for all military pilot selection in the United Kingdom since 1997.

A validation of a test battery via a meta-analysis of five studies between 1955 and 1998 used by the Norwegian Air Force was conducted by Martinussen and Torjussen (1998). For admission into the Norwegian Air Force, pilots must pass three stages of selection. After having completed at least twelve years of education, candidates are admitted to the first step in the selection process. In the first step candidates are given several paper-and-pencil tests designed to measure cognitive abilities, and they are also required to pass a physical examination. Candidates who pass this first step then move on to the second step during which they are subjected to cognitive and psychomotor tests. The tests are grouped into four different categories: general intelligence, technical comprehension and spatial ability,

simultaneous capacity, and orientation ability. Those who pass the second stage are given a defense mechanism test and are then interviewed by a psychologist. The last step is a panel interview and a final medical examination. The meta-analysis found that the best predictors of pilot performance during basic flight training were instrument comprehension ($r = 0.29$), mechanical principles ($r = 0.23$), and aviation information ($r = 0.22$).

Woycheshin (2002) conducted a validation study focusing on the Canadian Automated Pilot Selection System (CAPSS). CAPSS is a computerized simulator of a single-engine light aircraft introduced in 1997. The study compared the CAPSS scores of 161 trainees to primary flight training results used to select pilot applicants. The correlation coefficients between CAPSS scores and course results were moderate (flight performance ratings, $r = 0.35$). There was also a moderate relationship between previous pilot experience and CAPSS scores ($r = 0.35$), and between previous experience and course results (flight performance ratings, $r = 0.55$).

Commercial cognitive/psychomotor ability studies.

Trankell (1959) studied a battery of tests used in the Scandinavian Airlines System (SAS). SAS test scores were correlated to success in training. Significant correlations were found for inductive intelligence ($r = 0.33$), verbal intelligence ($r = 0.28$), and mechanical comprehension ($r = 0.21$).

Lufthansa German Airlines has used pilot selection tests developed by the Department of Aviation and Space Psychology of the German Aerospace Research Establishment for more than 35 years (Gnan, Flynn, & King, 1995). Lufthansa uses an ab-initio pilot program to select most of their student pilots. These ab-initio pilots are trained at pilot schools in Bremen, Germany and in Phoenix, Arizona. Every pilot applicant is required to complete the German university entrance level (Abitur) and cannot be older than 27. The selection process includes two steps. The first step is covered over two days in which all pilot applicants must take group-administered performance and personality tests that measure: Technical knowledge and comprehension; mathematics; concentration ability; perception speed; memory; spatial orientation; and, temperament characteristics. This first step usually reduces the number of pilot applicants by 70-75%. Those who successfully pass the first step continue on to the second step, which is conducted over three days. During the first day, psychomotor coordination and multiple task capacity tests are administered, an interview in front of the selection board is conducted on the second day, and finally a medical examination is performed on the third day. At the conclusion of the second step, only about 10% of the original group is accepted. Hörmann and Maschke reported that by using these methods Lufthansa German Airlines has reduced the attrition rate of ab-initio pilot training down to 3% (Hörmann & Maschke, 1996).

Personality/Job sample studies.

Although most pilot selection studies have focused on assessing cognitive abilities, some studies have attempted to assess personality and job sample measures. In samples of United States Navy pilot candidates that ranged from 1,818 to 2,356, Fiske (1947) reported correlations between the Biographical Inventory (a list of activities and events in a person's life history) and pass/fail scores ranged from $r = 0.30$ to 0.35 . Similarly, Taylor, Murray, Ellison, and Majesty (1971) developed two biographical inventories and an activities index that was administered to 645 USAF pilot students scheduled for training. Correlations between the biographical inventories and the activities index ranged from $r = 0.22$ to 0.32 .

Structured interviews have also been used in pilot selection studies. Walters, Miller, & Ree (1993) investigated a structured interview for the selection of USAF pilots. A total of 223 pilot trainees were interviewed. The interview yielded ratings of subject attributes which included: Educational background;

self-confidence and leadership; flying motivation; success in training; and, success in flying various classes of aircraft. The answers to each question were rated on a 5 point scale from poor (1) to good (5). Interview scores were then compared to AFOQT and BAT scores. The structured interview was found to be valid but lacked incremental validity.

Hörmann and Maschke (1996) investigated the validity of a personality questionnaire, the Temperament Structure Scales (TSS), for the prediction of job success of airline pilots compared to validities of a simulator check flight and flying experience data. The TSS is a multidimensional personality questionnaire with eight scales: Extraversion; dominance; emotional instability; aggressiveness; empathy; achievement motivation; rigidity; and, vitality. A total of 274 applicants were graded on a simulator check flight. Pilots were classified as performing at standard or below standard levels after three years of employment with the company. A multiple-regression model showed that job success could be predicted with 73.8% accuracy based on simulator check-flight scores and flying experience. By adding the TSS to the regression equation, job success could be predicted with 79.3% accuracy.

Job sample measures have been studied in previous pilot selection studies. In most cases student pilots were assessed in simulators or light-planes. Signori (1949) studied the Royal Canadian Air Force pilot selection methods during World War II. A total of 366 pilot candidates were tested through check-rides after seven and 11 hours of training were completed. Flight instructor scores were compared to success in flight training and yielded correlations of $r = 0.44$ (7 hours) and $r = 0.39$ (11 hours). In a study of light-plane training among U.S. Navy personnel, Ambler and Waters (1967) found that 15% of students who received light-plane training failed in pilot training, while 30% of those who had not received light-plane training failed.

Perceptual Speed

The aptitude of perceptual speed is a highly valid predictor of job performance (Mount, Oh, & Burns, 2008). Perceptual speed tests measure two factors: speed of processing information and the ability to focus attention. A significant difference between perceptual speed and other aptitudes is that perceptual speed tests are “speed” tests while other aptitude tests are “power” tests. Perceptual speed test questions are fairly simple which means that all subjects should achieve perfect scores if given an ample amount of time. This means that errors on perceptual speed tests are probably a result of the inability to process the information quickly and/or the inability to focus attention to perceive the detailed stimuli.

Numerous studies have demonstrated the validity of perceptual speed as a predictor of job performance. Ghiselli (1966) found that perceptual accuracy predicted overall performance in clerical jobs. Pearlman, Schmidt, and Hunter (1980) found that perceptual speed tests predicted clerical job performance better than motor ability or spatial/mechanical ability. Ackerman and Cianciolo (2000) found that perceptual speed tests account for incremental validity in task performance over general mental ability. Mount et al. (2008) administered the Name Finding Test (NFT) to warehouse workers which measures perceptual speed. The NFT contains 32 different names and the subjects were to observe a name and remember the name well enough to identify it from a group of four names spelled in a similar manner. This test is highly relevant to warehouse workers because they must load and unload trucks quickly while reading invoices and packing lists that contain names of food items and the names of restaurants. This study found that the number of correct answers on the NFT was shown to predict task performance.

METHODOLOGY

The purpose of this study was to validate the Table Reading Test as an assessment tool to indicate potential success in flight training. Specifically, subject TRT scores were compared to the number of flight hours flown prior to each subject successfully flying solo, the number of flight hours accumulated prior to each subject successfully earning a private pilot's certificate, and the individual's most recent GPA. Analyses using Microsoft Excel were conducted to determine the correlations between the TRT and performance scores.

Subjects

Subjects were freshman enrolled in an aviation program at a university in the southwest. The test was administered between the Fall 2005 and Fall 2008 semesters, inclusive. While 277 students completed the TRT over the four-year period, criteria performance data could be collected for only 116 subjects. Valid flight criterion data were not available for some subjects for several reasons: some data collected from participants appeared to be estimated due to the non-use of a decimal place in reporting flight hours; student flight records are only required to be kept on file for the most recent two years, some students may have graduated before their data could be collected; students may have changed majors or transferred to another school before all the data could be collected.

Test Instrument

The TRT is a test of perceptual speed consisting of 50 multiple-choice (a-e) questions. Subjects are given 9 minutes to complete the test, however the test is designed so that test takers should not be able to complete it within the allotted time. Subjects are given a table sheet, a test booklet with the 50 questions, and a sheet to record their answers.

The table sheet is made up of X and Y values. The X values, or column values, scale across horizontally and range from -17 on the left side of the page to +17 on the right side of the page. The table also consists of Y values, or row values, that scale vertically and range from +17 to -17 from top to bottom. For each question, subjects are given an X and Y value which corresponds to a number on the table sheet. The student's task is to determine the number where the column and row intersect and select the correct number among the five answer options for each question. Subjects are not allowed to use a straight edge to assist in locating the values. Subjects are not penalized for incorrect answers. For example, a subject who attempts 30 questions and selects 30 correct answers earns the same score as someone who attempts 50 questions and selects 30 correct answers.

Test Administration

The TRT was administered to the subject pool over the four-year period. The instructor of the 100-level introductory aviation course administered the TRT at the beginning of each semester. At the start of a class period, the instructor would explain to the students the purpose of the test, and its role in this longitudinal investigation. Participation was voluntary. Students who participated were not compensated, and students choosing to not participate were not penalized. All materials, including test booklet, demographic/answer sheet, and table sheet would then be distributed to participants. They were instructed to not open the test booklet until told to do so. After all materials were distributed, the instructor would provide the participants three minutes to review the test directions and sample problems. At the end of three minutes, the instructor would tell them that they would then have nine minutes to work through the test. At the expiration of the nine minutes, the instructor would ask the participants to stop, and pass all

test materials forward. After the class period, data from the answer sheets was input into a Microsoft Excel spreadsheet, and sent electronically to the test developer for scoring.

Flight performance data were obtained from both paper, computer-based student flight records, as well as from personal requests sent out to all current flight students. The majority of the Time-to-Solo data was collected from student flight records on file at the school, along with the Education Training Administration (ETA) program, a recently-introduced (2008) computerized logbook. The remaining Time-to-Solo data was collected from surveys that students completed. All of the Time-to-Private (certificate) data was collected from the surveys.

Academic data was obtained from university administrative personnel. GPAs were provided for each student for the Fall semesters from 2005-2008. In some cases there were students who had more than one GPA on file (Fall 2005, Fall 2006, etc.). Each subject's GPA directly after they soloed and/or earned their private pilot's certificate was not available, so the last recorded GPA for each student was used. The TRT scores were then individually matched to each subject's recorded flight and GPA data.

Data Analysis

Once the TRT was scored, all TRT scores and criterion data were tabulated and organized in Microsoft Excel. Correlation analyses were then conducted to determine whether there was a relationship between TRT scores to the number of flight hours accumulated prior to successfully flying solo, the number of flight hours accumulated prior to successfully earning their private pilot's certificate, and the most recent GPA.

When performing the analyses, only subjects that had a TRT score on record were considered. The researcher elected to adhere to the statistical community's standard of defining significant results if the test's p-value is less than 5%. Hence, any p-values greater than or equal to 5% would be considered to indicate results that are not significant.

Because this present study seeks to validate the TRT, correlation analyses were conducted in order to investigate the correlative relationship between TRT scores and the performance criterion. Regression analyses were not used in this study because the researcher did not want to give any indication of cause and effect; that is, a certain TRT score will result in a confidence interval of flight hours to soloing and/or obtaining a private pilot's certificate.

RESULTS

The results of the three correlation analyses follow. A description of the results of each individual analysis is included in detail.

Results

Three correlation analyses were conducted to determine the relationship of TRT scores to the flight and academic performance criterion. TRT scores were compared to Time-to-Solo data, Time-to-Private data, and subject GPA. All analyses yielded significant results. Table 1 presents a summary of the data and results for each analysis.

Table 1: TRT Scores and Flight/Academic Performance Criterion Results

Analysis	Performance Criterion	n	Correlation Coefficient	$p <$
1	Time-to-Solo	99	-0.228	0.024
2	Time-to-Private	17	-0.754	0.001
3	GPA	116	0.283	0.002

Note. All p -values satisfy the aforementioned $p < 0.05$ rule.

The first analysis focused on the number of flight hours accumulated by each subject prior to each successfully flying solo. The results were found to be significant, $F(1, 97) = 5.295$ and $p < 0.024$. The second analysis considered the number of flight hours accumulated by each subject prior to each successfully earning a private pilot's certificate. The results were also found to be significant, $F(1, 15) = 19.837$ and $p < 0.001$. The third analysis focused on the last recorded student GPA. The results were likewise found to be significant, $F(1, 114) = 9.973$ and $p < 0.002$.

Summary of Results

Three significant correlations ($p < 0.05$) were found between the TRT scores and the performance criteria data:: Subject Time-to-Solo, Time-to-Private, and GPA. The data indicate that the strongest correlation is between the TRT scores and number of hours each subject accumulated prior to earning his or her private pilot's certificate ($r = -0.754$, $p < 0.001$). This is particularly important insofar as resources for training new pilots are becoming increasingly more scarce, both for students and schools. Hence, the TRT could be a means to better screen students before they fully embark down the path of either becoming a licensed private pilot, or needlessly pursuing a costly dream.

DISCUSSION

This section includes discussion and explanation of the correlation analyses performed. In addition, future recommendations for administering the test, collecting data, as well as other procedures are included along with a discussion on how these results may be utilized. This section concludes with recommendations for future research.

Significant Findings

The result of the first analysis (time-to-solo) was $r^2 = 0.052$ and $p < 0.001$. This implies that 5.2% of the variation in the TRT scores due to the different subject abilities is in common with a like amount of variation in subject hours prior to soloing. This is not the same as cause and effect, as there is no evidence that this exists between these metrics. Nonetheless, the metrics do share a common data source, the student subjects and the r^2 values are an indication of a common source of performance variation.

The result of the second analysis (time-to-private) was $r^2 = 0.569$ and $p < 0.001$. This implies that 56.9% of the variation in the TRT scores due to the different subject abilities is in common with a like amount of variation in subject hours prior to earning his or her private pilot certificates. In other words, nearly 57% of the variability in the number of hours prior to earning a private pilot certificate is attributable to the variability of TRT scores. Again, note that this is not the same as cause and effect.

The result of the third analysis (GPA) was $r^2 = 0.080$ and $p < 0.001$. This implies that 8.0% of the variation in the TRT scores due to the different subject abilities is in common with a like amount of variation in subject GPAs. This means that 8% of the variability in subject GPAs is attributable to the variability of TRT scores.

Results Applications

As stated previously, all three correlation analyses yielded significant results, most notably TRT vs. Time-to-Private. These results seem to indicate that the TRT could be used to recruit middle school and high school students to consider a career as a professional pilot. While many students have never considered becoming a pilot, they could be encouraged to pursue a career as a pilot, if they perform well on the test. Conversely, a student's score may also indicate that they may have some difficulty in successfully completing pilot training, so they may then wish to consider other career options in and out of the aviation industry.

Future Considerations

With so many factors and variables that affect the outcome of each analysis it is important to limit their effects in order to get a clear understanding of the results of the test and how those results can help predict future pilot performance. There are a few things that should be considered in the future when conducting a study similar to this, one of which is the data collection. It was clear when collecting the data from student flight records that flight data, in particular number of hours to solo, was inconsistent. This can be resolved by possibly acquiring flight data from a computerized logbook that students use (the ETA program), or personally verifying recorded flight hours in each subject's logbook.

In the future, students upon successfully soloing and earning their private pilot's certificate should record the exact number of hours prior to completing each task as well as report their current GPA. They should also take the TRT again in order to compare their previous test score to their scores after completing each task. By requiring students to take the TRT again, researchers can further assess the reliability of the test.

After interacting with some of the subjects, it should be noted that some flight times might have been affected due to a change in flight instructors. Some subjects reported that they had two and even three flight instructors, which may have caused them to fly additional hours with their new flight instructor to get acclimated with each other. Future studies should note when subjects have more than one flight instructor when analyzing flight times.

Future researchers should conduct a predictive study to find the degree to which TRT scores can predict future pilot performance. This study only considered flight performance up to the private pilot certificate rating. Future researchers should also focus studies up to the flight instrument rating, where pilots are required to read flight instruments in the cockpit, which is a more visually demanding task for pilots.

This study as well as most studies in the past focused on the number of correct answers. Future researchers may wish to focus on analyzing the number of incorrect answers as well as a looking at scores as a percentage of right/wrong answers vs. the total number of questions attempted. Mount et al. (2008) suggest that the number of wrong answers on perceptual speed tests may predict inattention or lack of rules compliance such as accidents, safety violations, and tardiness.

Summary

With rising flight training costs, pilot selection studies are needed to select individuals who have the greatest potential to become successful pilots. In the past the U.S. military, various airlines, and other aviation entities have tried to determine and assess the qualities required to become a successful pilot. In an attempt to determine these qualities, most pilot selection studies have focused on cognitive abilities while others focused on personality and job sample measures. The focus of this study was to assess and validate the Table Reading Test, which was designed to measure perceptual speed. With cockpits becoming more sophisticated and visually interactive, perceptual speed is becoming a more critical skill for pilots.

This correlative study has been able to demonstrate some degree of validity with regards to the TRT. This study shows that higher TRT scores are associated with better flight training results, most notably the number of hours accumulated prior to earning a private pilot certificate. The TRT can help determine the success or flight performance of a student who may be unsure in their pursuit in becoming a pilot. It can also encourage students to pursue becoming a pilot who may not have even considered becoming a pilot. A longer, more in-depth predictive study might further validate the test and show it to be a strong predictor of future pilot performance.

REFERENCES

- Ackerman, P., Cianciolo, A. (2000). Cognitive, Perceptual-speed, and Psychomotor Determinants of Individual Differences During Skill Acquisition. *Journal of Experimental Psychology: Applied*, 6(4), 259-290.
- Ambler, R. K., & Waters, L. K. (1967). *The Value of an NROTC Flight Indoctrination Program to Naval Aviation Training*. Pensacola, FL: Naval Aerospace Medical Center.
- Bailey, M., & Woodhead, R. (1996). Current Status and Future Developments of RAF Aircrew Selection. *Selection and Training Advances in Aviation: AGARD Conference Proceedings 588* (81-89). Prague, Czech Republic: Advisory Group for Aerospace Research & Development.
- Bair, J. T., Lockman, R. F., & Martoccia, C. T. (1956). Validity and Factor analysis of Naval Air Training Predictor and Criterion Measures. *Journal of Applied Psychology*, 40(4), 213-219.
- Carmines, E. G., & Zeller, R. A. (1979). *Reliability and Validity Assessment*. Beverly Hills, CA: Sage Publications.
- Carretta, T., & Ree, M. (1995). Air Force Officer Qualifying Test Validity for Predicting Pilot Training Performance. *Journal of Business and Psychology*, 9(4), 379-388.
- Carretta, T., & Ree, M. (1996). Central Role of g in Military Pilot Selection. *The International Journal of Aviation Psychology*, 6(2), 111-123.
- Carretta, T., & Ree, M. (1998). Computerized Testing in the United States Air Force. *International Journal of Selection and Assessment*, 6(2), 82-89.
- Carretta, T., & Ree, M. (2000). *Pilot Selection Methods*. (USAF Research Laboratory No. AFRL-HE-WP-TR-2000-0116). Wright-Patterson AFB, OH. Retrieved on January 27, 2009 from <http://www.dtic.mil/cgibin/GetTRDoc?AD=ADA384132 &Location=U2&doc=GetTRDoc.pdf>.
- Damos, D. (1996). Pilot Selection Batteries: Shortcomings and Perspectives. *The International Journal of Aviation Psychology*, 6(2), 199-209.
- Damos, D. (2004). *Table Reading Test Version 2*. Los Angeles, CA: Damos Aviation Services.
- Di Iorio, C. K. (2005). *Measurement in Health Behavior: Methods for Research and Education* (1st ed.). San Francisco: Jossey-Bass.
- Fiske, D. W. (1947). Validation of Naval Aviation Cadet Selection Tests Against Training Criteria. *Journal of Applied Psychology*, 31(6), 601-614.
- Ghiselli, E. (1966). *The Validity of Occupational Aptitude Tests*. New York: Wiley.
- Gnan, M., Flynn, C., & King, R. (1995). (AL/AO-TR-1995-0003). Brooks Air Force Base, TX. Retrieved on July 25, 2009, from <http://www.dtic.mil/cgibin/GetTRDoc?AD=ADA293779 &Location=U2&doc=GetTRDoc.pdf>.

- Goeters, K. (Ed.). (1998). *Aviation Psychology: A Science and a Profession*. Hants, England; Brookfield, Vt.: Ashgate Publishing Limited.
- Goeters, K. (Ed.). (2004). *Aviation Psychology: Practice and Research*. Hampshire, England; Burlington, Vt.: Ashgate Publishing Limited.
- Griffin, G., & Koonce, J. (1996). Review of Psychomotor Skills in Pilot Selection Research of the U.S. Military Services. *The International Journal of Aviation Psychology*, 6(2), 125-147.
- Henmon, V. A. C. (1919). Air Service Tests of Aptitude for Flying. *Journal of Applied Psychology*, 3(2), 103-109.
- Hörmann, H. & Maschke, P. (1996). On the Relation Between Personality and Job Performance of Airline Pilots. *The International Journal of Aviation Psychology*, 6(2), 171-178.
- Hunter, D. R., & Burke, E. F. (1995). *Handbook of Pilot Selection*. Aldershot, Hants, England; Burlington, Vt.: Ashgate Publishing Company.
- Martinussen, M. & Torjussen, T. (1998). Pilot Selection in the Norwegian Air Force: A Validation and Meta-Analysis of the Test Battery. *The International Journal of Aviation Psychology*, 8(1), 33-45.
- Messick, S. (1995). Validity of Psychological Assessment: Validation of Inferences From Persons' Responses and Performance as Scientific Inquiry Into Score Meaning. *American Psychologist*, 50(9), 741-749.
- Mount, M., Oh, I., & Burns, M. (2008). Incremental Validity of Perceptual Speed and Accuracy Over General Mental Ability. *Personnel Psychology*, 61, 113-139.
- North, R. A., & Griffin, G. R. (1977). *Aviator Selection 1919-1977* (Special Report 77-2). Pensacola, FL: Naval Aerospace Medical Research Laboratory.
- Pearlman, K., Schmidt, F., & Hunter, J. (1980). Validity Generalization Results for Tests Used to Predict Job Proficiency and Training Success in Clerical Operations. *Journal of Applied Psychology*, 78(4), 518-524.
- Rosenberg, M. (1965). *Society and the Adolescent Self-Image*. Princeton, NJ: Princeton University Press.
- Signori, E. I. (1949). The Arnprior Experiment: A Study of World War II Pilot Selection Procedures in the RCAF and RAF. *Canadian Journal of Psychology*, 3(3), 136-150.
- Taylor, C. W., Murray, S. L., Ellison, R. L., & Majesty, M. S. (1971). *Development of Motivation Assessment Techniques for Air Force Officer Training and Education Programs: Motivation for Pilot Training*. (AFHRL-TR-71-21). Brooks Air Force Base, TX: Professional Education Division, Air Force Human Resources Laboratory.
- Thorndike, R. L. (1949). *Personnel Selection: Test and Measurement Techniques*. New York: J. Wiley.
- Trankell, A. (1959). The Psychologist as an Instrument of Prediction. *Journal of Applied Psychology*, 43(3), 170-175.

- Walters, L. C., Miller, M. R., & Ree, M. J. (1993). Structured Interviews for Pilot Selection: No Incremental Validity. *The International Journal of Aviation Psychology*, 3(1), 25-38.
- Want, R. L. (1962). The Validity of Tests in the Selection of Air Force Pilots. *Australian Journal of Psychology*, 14(2), 133-139.
- Woycheshin, D. (2002). Validation of the Canadian Automated Pilot Selection System (CAPSS) Against Primary Flying Training Results. *Canadian Journal of Behavioural Science*, 34(2), 84-89.