Pilot Selection Systems Help Predict Performance

Research shows that structured selection systems can identify applicants who possess the knowledge, skills, abilities and personality traits most valued by a particular aircraft operator and who will succeed as pilots in that operator's line operations.

Diane L. Damos, Ph.D.

Air carriers, on-demand operators and corporate flight departments use various methods to identify the pilots they want to employ among applicants who meet minimum requirements of certification and experience (see "Hiring Procedures Differ Among Airlines," page 2). Recent research has shown that structured pilot selection systems are most effective in helping aircraft operators to identify pilots who match their requirements, compared with casual selection systems.

In casual selection systems, the interview often is unstructured, and the interviewer may not be trained in interview techniques. Typically, there is no systematic method of combining information from the interview, the flight-skills test and the application form; a hiring decision, therefore, is based solely on a manager's judgment, which sometimes can be influenced by a variety of subjective factors. Casual pilot selection systems typically have not been developed by specialists with appropriate expertise, and often there is no historical documentation of how and why they came to exist. These systems also may lack explicit hiring standards, and the manager often has only a general idea of how to determine which applicants would best match the company's requirements.

Casual systems typically fail to accomplish three major goals common among pilot selection systems:

- To identify the pilots who best match the operator's requirements;
- To accomplish this identification in a cost-efficient manner; and,
- To be legally defensible if the process or selection decisions are challenged.

In contrast, structured pilot selection systems — which typically include a consistent set of written and/or computer-based tests of knowledge, skills and abilities (KSAs); personality assessments; interviews; and simulator assessments (check rides) — achieve these goals in the following ways:

- Developing the system requires management to analyze the KSAs and personality traits required by the aircraft operator. This development process often reveals many different assumptions and disagreements among individuals directly involved in the hiring process. Developing the system requires management to evaluate these assumptions and to confront differences in opinion about which KSAs and personality traits are required;
- Hiring decisions are made through the use of a decision aid — a statistical process that combines scores from all tests administered as part of the pilot selection process to predict how well the applicant will perform as a pilot in the company. The decision aid does not eliminate the human decision maker but rather reduces the subjectivity of the hiring process;

Continued on page 3

Hiring Procedures Differ Among Airlines

Some air carriers determine whether a pilot will be hired after a single interview; others base their hiring decisions not only on interviews but also on the pilot's performance on a number of tests.

Capt. Kit Darby, owner and president of Air Inc., an Atlanta, Georgia, U.S., career information resource company that provides services to pilots seeking airline jobs and administers tests for smaller U.S. airlines, said that the hiring process is "consistently inconsistent."[1]

"There are major airlines that don't do anything but [conduct] a 45-minute interview," said Darby, also a Boeing 767 captain with United Airlines. "Others have more rigorous testing."

Several major U.S. airlines administer complete sets of written tests or computer-based tests, more than half administer psychological tests, most administer simulator evaluations, and almost all administer aeronautical knowledge tests, he said. A few airlines have eliminated simulator evaluations in favor of cognitive tests (typically computer-based tests or paper-and-pencil tests intended to measure an individual's reasoning ability), he said.

Pilot selection systems are most effective when they include thorough and consistent testing, Darby said. As an example, he cited one airline that added cognitive tests to its pilot selection system without telling its ground instructors that the system had been changed. The instructors became aware of a change, however, because they observed that, compared with the previous students as a whole, the new students were quicker to understand concepts being taught in the classroom.

Darby's company monitors the hiring activities of about 200 U.S. airlines and projects that they will hire about 7,000 pilots in 2003, compared with 6,000 pilots in 2002, 12,000 pilots in 2000 and 19,000 pilots in 1999, which was the sixth consecutive annual record. Of the 7,000 pilots expected to be hired this year, about 500 pilots will be hired by major airlines, he said. About 7,500 U.S. pilots have been furloughed.

Capt. Charles Hogeman, vice chairman of the Air Line Pilots Association, International (ALPA) Pilot Training Council, said that he has "mixed feelings" about the elimination of simulator evaluations.[2]

"Simulator evaluations for hiring pilots are just snapshots of a pilot's ability," Hogeman said. "There are a number of enterprises that prepare applicants for these evaluations — teach the test — so I'm not sure they are a real gauge of the applicant's real flying ability. However, if applicants are evaluated evenly and fairly, a simulator evaluation may identify applicants who may not have sufficient skills to adapt to training in jet transports."

Hogeman said that he is "skeptical" about the extent to which cognitive tests can aid in the selection of new pilots but that "valuable insight to applicant thought processes and attitudes toward flying airplanes can be discovered through a well-managed personnel interview."

Sharon Jones, a flight operations recruiter in the human resources department of Comair, said that her airline is one of those that has eliminated the simulator evaluation because "it may not give an accurate reading of the applicant's ability to fly the airplane."[3]

The pilot selection process at Comair — which has 1,600 pilots and expects to add 388 pilots in 2003 — includes interviews, a short psychological test and a computer-based cognitive test.

"Cognitive screening gives us a better opportunity to test their skills in a certain way," said Nick Miller, Comair manager of media relations.

Jones said that Comair's selection process has "always been well-defined. Years of refining that process have proven that's what works best."

Capt. Christian Magnusson of Scandinavian Airlines System (SAS) said that for the past two years, the airline has been working to determine "how to recruit a new type of pilot into SAS — not because the recruiting profile we use is wrong but because it needs to be updated."[4]

Magnusson, manager of the project, said that the work has included a task analysis "to see what a pilot in SAS really does today," an assessment of the requirements of SAS aircraft and a review of the abilities of young job applicants.

"We're looking more into the non-technical skills required of pilots — communicating, understanding other people's points of view, managing a system involving an airplane and people," he said. "Pilots in the future need to have basic flying skills but that isn't enough. They need to fly an aircraft and also manage the system."

Previous pilot applicants have undergone three days of tests and interviews, including one day of psychosocial tests and written tests, one day of simulator evaluations and one day of interviews. New tests are being developed and should be in place for applicants in late 2004 or in 2005 to assess leadership and management skills, Magnusson said; those tests will involve evaluations by psychologists and pilots of applicants' performances on non-written problem-solving tests.

Representatives of several air carriers contacted for this article did not respond to questions about their pilot selection

procedures. Some air carriers, however, discuss their hiring processes — in varying detail — in printed information or on their Internet sites.

Some airlines tell prospective applicants what to expect during the selection process, describe minimum requirements for pilot applicants and/or provide considerable detail on their selection systems.

For example, British Airways describes the personality traits it values in its pilots: "leadership, determination, reliability, high personal standards, motivation, flexibility, well-developed customer-service skills and teamwork."[5]

Qantas Airways describes its tests — four aptitude tests designed to assess verbal reasoning, numerical reasoning, diagrammatic abilities and spatial abilities — and questionnaires — two questionnaires dealing with motivation and personality. Qantas also provides samples of the types of questions that will be included on each test and questionnaire, provides suggestions for improving test performance and offers the following advice:[6]

> Don't be discouraged if you found the questions difficult. There are several things you can do to improve your performance. ... Read newspapers, reports, business journals; do verbal problem-solving exercises [e.g., crossword puzzles]; read financial reports in newspapers; study tables of data; practice your mental arithmetic; solve puzzles in newspapers and magazines involving diagrams; play games involving sequences or strategies (e.g., chess ...); look at plans and [do-it-yourself] manuals; make up patterns or designs; [and] imagine how various objects look from different angles.♦

— FSF Editorial Staff

Notes

1. Darby, Kit. Telephone interview by Werfelman, Linda. Alexandria, Virginia, U.S. Jan. 28, 2003. Flight Safety Foundation, Alexandria, Virginia, U.S.

2. Hogeman, Charles. E-mail communication with Werfelman, Linda. Alexandria, Virginia, U.S. Feb. 5, 2003. Flight Safety Foundation, Alexandria, Virginia, U.S.

3. Jones, Sharon; Miller, Nick. Telephone interview by Werfelman, Linda. Alexandria, Virginia, U.S. Feb. 7, 2003. Flight Safety Foundation, Alexandria, Virginia, U.S.

4. Magnusson, Christian. Telephone interview by Werfelman, Linda. Alexandria, Virginia, U.S. Feb. 10, 2003. Flight Safety Foundation, Alexandria, Virginia, U.S.

5. British Airways. *Direct Entry Pilot Scheme: The Qualities We Seek and How We Develop Them.* <www.britishairwaysjobs.com/cc/pilots/deps/qualities.jsp>. Jan. 31, 2003.

6. Qantas. *Qantas Pilot Psychometric Test Practice Leaflet.* <www.qantas.com.au/info/about/employment/pilots>. Feb. 3, 2003.

- System results are easy to evaluate quantitatively. All pilot selection systems are, in some sense, methods of prediction (i.e., they predict how well a job applicant will perform as a pilot in the company). The predictive validity of the pilot selection system can be tracked over time. (Predictive validity is the determination of the extent to which the scores on a test predict actual performance.) A decrease in predictive validity may indicate a need to revise parts of the pilot selection system, to look for new sources of pilot applicants or to change the initial training of newly hired pilots. This measurement of results also allows the cost-effectiveness of the pilot selection system to be determined relatively easily; and,

- The system relies on documentation. In countries with laws prohibiting discrimination in hiring — for example, the United States and South Africa — documentation can be used to defend the company before courts and regulatory bodies. The documentation also provides important "corporate memory" for the human resources department and the flight operations department.

The statistical methods used in a structured pilot selection system require that, when an aircraft operator begins to hire, approximately 30 new pilots begin training each year. (The number of pilots who complete training is not relevant; nor is the total number of pilots employed by the operator.) If an operator does not hire anyone for five years and then hires approximately 30 new pilots in the sixth year, this system can be used.

Some air carriers and other aircraft operators have employees with the specialized skills required to manage a structured pilot selection system, such as technical knowledge in employment law and the mathematics of selection, including knowledge of scientific literature involving pilot selection. Aircraft operators who do not have employees with those skills sometimes hire consultants to perform the required pilot selection tasks.

Selection Systems Differ From Screening Systems

In informal discussions, screening systems frequently are confused with selection systems. A screening system is designed to eliminate from consideration for employment any applicant who does not meet the minimum requirements for

hiring. For example, an air carrier might require a pilot to have 1,000 flight hours to be considered for a job. Nevertheless, the air carrier might receive applications from pilots with fewer than 1,000 flight hours. A pilot-screening system would eliminate those pilots.

A pilot-screening system also might require background checks to identify job applicants with criminal records or histories of alcohol abuse or drug abuse and to ensure that applicants have authentic credentials to show that they have the education, certification and experience that they say they have.

A selection system is designed to identify the best job applicants.

Selection systems use "select-in" processes to identify applicants who best meet operator requirements and to bring them into the company; screening systems use "select-out" processes to eliminate unacceptable applicants from consideration.

An aircraft operator's human resources department should allow a pilot applicant to begin the selection process only if he or she has successfully completed the screening process.

A structured pilot selection system consists of the following five major elements:[1]

- The criterion is the measure of success in training or success on the job. The criterion — which establishes what type of individual will be hired — is the most important element of the selection system because it represents the aircraft operator's definition of the KSAs and personality traits required in the company's pilots.[2] An inappropriate criterion might result in the hiring of new pilots who do not meet the operator's expectations. Some aircraft operators ignore this element and say only that they attempt to select "good pilots" rather than specifying exactly what KSAs and personality traits are required by the operator, such as a person who will complete initial training within the "standard footprint" (the typical amount of time and resources allowed for training), will score at least 90 points out of 100 points on the final simulator assessment and will arrive on time 98 percent of the time for his or her flights). Those specifications are criterion measures;

- Testing may include written tests or computer-based tests of intelligence, personality traits, motor skills and information-processing ability. Testing also may include interviews to assess knowledge and experience and assessments of flying skills, such as simulator assessments. The battery (group or set) of tests to be used should be identified relatively late in the development of a structured pilot selection system, after the criterion has been established, and the tests should complement the criterion;

- Models determine when qualified applicants will be eliminated from the group of those being considered for pilot jobs. A structured pilot selection system is designed around one of three models — the single-hurdle model, the multiple-hurdle model or the progressive model. The choice of model affects the cost per applicant of the pilot selection process.

  The simplest model is the single-hurdle model (Figure 1). An aircraft operator using this model administers all tests to all candidates for the job before making any hiring decisions. Thus, every pilot applicant receives all of the written/computerized tests, all of the interviews and a simulator evaluation. This model is used rarely because of the per-applicant costs.

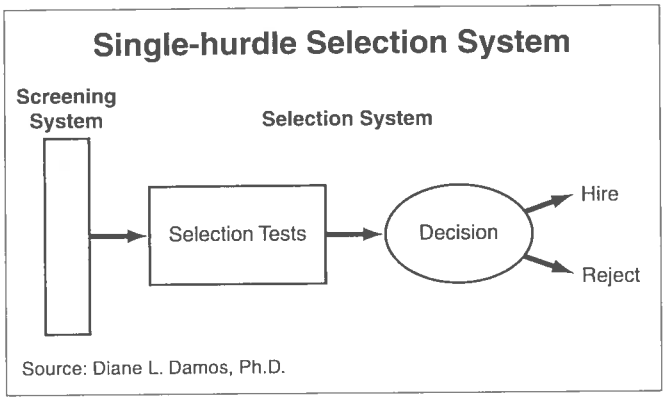

Figure 1

A pilot selection system used by Scandinavian Airlines System (SAS)[3] was the only example of the single-hurdle model found to have been analyzed in scientific literature. A 1959 report said that in this SAS pilot selection system, applicants were administered written tests and other tests of personality, motor skills, intelligence and "timesharing" (simultaneous performance of multiple tasks). The tests were administered independently by two psychologists or three psychologists. After all tests had been administered and scored, the psychologists met, reached an agreement on each job applicant and sent their hiring recommendations to SAS management for a final decision. The process was costly and time-consuming.

The costs of the single-hurdle model may be justified primarily when applicants must be transported over long distances. For example, if an Asian air carrier recruits its pilots in North America and Europe, the costs involved in transporting applicants to Asia may be substantial, and the air carrier may want all applicants to complete all selection tests at one time so that the complete results can be used in hiring decisions.

Most structured pilot selection systems use a multi-hurdle model (Figure 2, page 5) that consists of at least two stages.[4,5] For example, when an aircraft operator

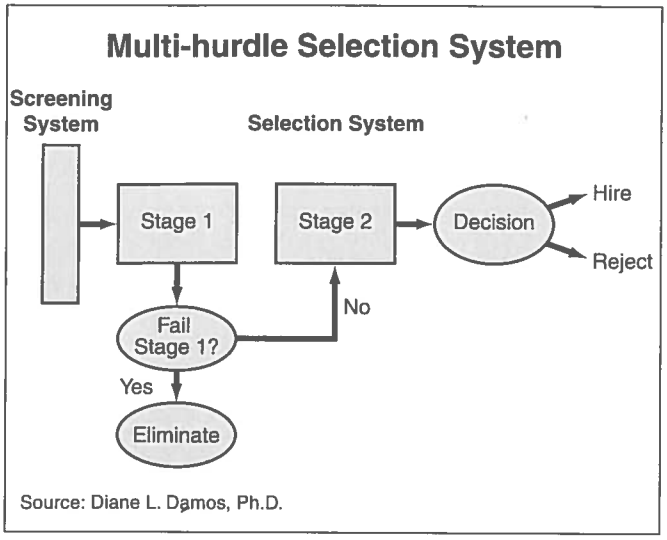

Figure 2

uses a multi-hurdle model with two stages, the selection tests are divided among the two stages, with the least expensive tests administered during the first stage and the most expensive tests administered during the second stage. All applicants complete the first-stage tests. The tests then are scored, and a substantial proportion of the lowest-ranking applicants are dropped from further consideration. The remaining applicants then complete the second-stage tests. The test results are analyzed with a decision aid, and a hire/reject decision is made for each applicant. The stages may occur successively on the same day, or they may be separated by substantial periods of time. Typically, the per-applicant costs for a structured pilot selection system with a multi-hurdle model are less than those for a single-hurdle model.

The third model is a progressive model, also known as a rolling model or a cascade model. This model is essentially a multi-stage model but differs on two major philosophical issues from the single-stage model and the multi-stage model. One major difference is that the pilot selection process continues during training, rather than ending at the time of the hiring decision.[6] The second difference is that the progressive model is designed to help ensure that the aircraft operator does not overlook an applicant who would perform well on the criterion measures; both the single-stage model and the multi-stage model are designed to minimize the possibility that an operator would hire an applicant who subsequently would perform poorly on the criterion measures. A progressive model is most useful when training is conducted over many months or years and when the applicant pool is small, compared with the number of available positions. Progressive models most often are used for ab initio systems because ab initio training typically requires several years;

- The decision aid uses a variety of statistical methods to relate scores on the tests to the criterion to produce an employment recommendation. The most commonly used statistical method is regression analysis, in which the selection scores on one or more tests are used to predict performance on a criterion; and,

- A feedback loop is used to monitor the predictive validity of a structured pilot selection system. All selection data for newly hired pilots should be recorded in a database. The feedback loop correlates the hired pilot's criterion measures with his or her scores on selection tests. Periodically, a manager calculates the predictive validity of the system. The predictive validity of any selection system decreases over time because of factors such as changes in the applicant pool, changes in the criterion or changes in the initial training. The feedback loop enables the aircraft operator to determine when the predictive validity is decreasing and to take appropriate countermeasures. In countries where employment litigation is common, the fact that a company has a feedback loop shows that the company has an intent to improve its pilot selection process and, therefore, provides an additional measure of protection against litigation.

Selection systems must be designed and must be administered to comply with all applicable regulations and employment laws, although complying with employment laws typically is not considered an element of a structured pilot selection system.

Selection-system Development Includes Five Steps

The first selection systems were developed by industrial/organizational psychologists more than 50 years ago.[7] Selection systems for pilots typically are developed according to the same principles as selection systems for other categories of employees.[8,9]

Aircraft operators should follow five steps in the development of a structured pilot selection system. The steps need not begin in the order given (i.e., step 2 may begin before step 1), and some steps may overlap, but all steps must be completed in order (i.e., step 1 should be completed first, then step 2, step 3, step 4 and step 5). The process and the outcome of each of the five steps also must be documented carefully for internal use and for defending against subsequent legal challenges.

Step 1 is conducting a job analysis to identify the KSAs and personality traits required to perform successfully as a pilot for a specific operator. The job analysis describes what the pilot does on the job, what the pilot must know to perform the job and what resources the pilot must use to perform the job successfully.

Much of the job analysis for a pilot involves task analysis, which is a detailed description of a pilot's work activities. An

aircraft operator may obtain a task analysis for its pilots by one of three methods:

- If the air carrier participates in an advanced qualification program (or other voluntary alternative to traditional regulatory requirements for pilot training), then the carrier already should have a task analysis of the pilot's job that was developed for training purposes. That task analysis can be modified for use in a structured pilot selection system. This is the best method to obtain a task analysis for the selection system because the analysis has been designed specifically for the operator's fleet and procedures;

- Some specialists (either in-house employees or outside consultants) in structured pilot selection systems have task analyses for pilots that can be modified for a specific aircraft operator. (Consultants may not have task analyses for the operator's aircraft, however.); and,

- An air carrier may conduct the task analysis itself. Performing a task analysis for a pilot job description can be a time-consuming process and can require substantial expertise, so this method of obtaining a task analysis is often beyond the in-house resources of many smaller aircraft operators.

The results of a task analysis rarely identify the personality traits that are required to perform well at a specific company; other techniques must be used to identify those traits. Usually, aircraft operators have strong opinions about the personality-trait requirements for the pilots they want to hire, and the most effective method to identify those requirements is to ask subject matter experts (SMEs), who typically are chief pilots and senior captains. These SMEs can identify the problems that the aircraft operator has experienced with its pilots (i.e., insufficient job skills, lack of dependability, poor leadership skills or poor stress management). They also can identify the personality traits of successful pilots, such as conscientiousness — typically one of the most highly valued personality traits for pilot applicants — and leadership abilities. This information may be obtained through interviews, surveys or small discussion groups, and after the information has been obtained, a specialist in structured pilot selection systems typically matches each of the successful-pilot traits with a generally recognized personality trait. After this step is completed, the aircraft operator will have a list of all of known KSAs and traits of a successful pilot within the company.

Step 2 is the identification of the tests (or types of tests) to be considered for inclusion in a selection system.

Identifying appropriate tests usually is a straightforward procedure for human resources specialists. Typically, a specialist in structured pilot selection systems recommends tests that measure the required KSAs and personality traits identified in step 1. If more than one test is available to measure a specific KSA or personality trait, the specialist typically makes a recommendation based on the cost per applicant, the time required to administer the test, the reliability of the test and the probable predictive validity of the test, based on the experience of other operators.

Sometimes there are no existing standardized tests to measure a specific KSA or personality trait. In such circumstances, the aircraft operator has two choices: develop a test to measure the KSA or trait or choose not to assess the KSA or trait. Test development should be pursued if the KSA or trait is considered very important for pilot success. Aircraft operators that do not employ individuals with expertise in test-development must hire specialists. Most commonly, tests cannot be administered to assess all KSAs and personality traits identified in step 1 because of time limitations; some tests must be omitted.

Step 3 requires the identification or development of performance measures to serve as the criteria for the pilot selection system. For example, one performance measure, such as the score on the check ride at the end of training, rarely assesses all of the important facets of a pilot's job for a particular aircraft operator.

Structured pilot selection systems can be designed to predict several attributes of the criterion. For example, an air carrier may want to predict measures of a pilot's performance during training, during probation and during operational flying. Many measures are generated during training that can be used in the criterion, such as time to complete training, the score on the final check ride or the score on the final oral examination. Computer-based training (CBT) programs also may generate a number of measures of a pilot applicant's progress during training. During probation, the score on the check ride administered at the end of probation and captains' assessments of an applicant are potential measures. Performance as an operational pilot may be reflected in check ride scores and the time required to complete recurrent training; CBT measures also may provide useful indicators of performance during recurrent training or upgrade training. Scores on annual check rides may be useful, and personnel records may include valuable measures that have been overlooked, such as the number of times during a 12-month period that a pilot arrives late for a flight.

Nevertheless, the following difficulties may occur in assessing pilot performance:

- Many common measures of performance in aviation are expressed as pass/fail. Such measures, which are referred to as "dichotomized" scores, typically are not good criterion measures, and accurate performance prediction is difficult.[10] Measures that use carefully developed scales are preferable; and,

- Occasionally, there are no performance measures that are appropriate for the criterion. For example, the aircraft

operator may have no measures of how well a pilot interacts with other crewmembers. In that event, the operator must develop performance measures. Nevertheless, development of performance measures requires specific technical knowledge and should be undertaken in consultation with appropriate specialists.

Step 4 requires administration of the tests to obtain predictive measures, to collect the criterion data, when available, and to identify the most cost-efficient tests.

This step begins the process of determining how well the chosen tests actually predict performance. Traditionally, in aviation, this step has been conducted by administering the new tests to a group of pilot applicants who have passed the screening procedure.[11] The scores on the new tests are not used in making an employment decision about the applicants; at this stage, the usefulness of the scores is unknown. Instead, management receives scores from the tests and/or interviews used previously and makes hiring decisions based on them. For example, if the aircraft operator has been using an interview and a simulator evaluation, these scores are used in making hiring decisions.

Newly hired pilots receive training, and those who complete training proceed to line flying. Scores on the criterion measures are collected at the appropriate time from each newly hired pilot. Statistical analyses would determine how accurately the scores on the new selection tests predict the criterion measures and how much they improve the predictive validity of hiring decisions. These analyses may show that some of the new tests are not cost-efficient and should be dropped from further consideration. Similarly, the analyses may show that some of the tests that the aircraft operator had been using are not cost-efficient and should be eliminated.

After identification of the tests that best predict the criterion, the resulting set of tests may be given to the next group of applicants. For this group, the scores from all of the tests are given to management. After the criterion data are collected, statistical analyses again are used to calculate the predictive validity and to help ensure that the predictive validity is sufficient for the operator's requirements.

Step 5 is the development of a monitoring system. Before further groups of applicants are tested, management should establish a monitoring system so that they will know whether the pilot selection system continues to predict the criterion measures established by the aircraft operator.

A typical monitoring system collects test scores and criterion scores, performs the statistical analysis and alerts management if the predictive validity of a structured pilot selection system decreases below a preset value. If a decrease occurs, the system notifies management that changes are required in the set of tests or in the analyses (decision aid). Management should identify reasons for the decrease (changes in the pool of pilot applicants, changes in training content or standards, etc.) and should take corrective action.

The monitoring system also documents the aircraft operator's intent and actions to ensure that a structured pilot selection system functions as intended and is improved periodically.

Selection Tests Assess Many Skills and Abilities

Many types of tests, including intelligence tests,[12] currently are used to assess KSAs and personality traits. Since World War II, scores on intelligence tests have been related demonstrably to performance in ab initio military pilot training.[13] Research from Qantas Airways[14] and SAS[15,16] shows that scores on these tests also provide a valid prediction of training performance and operational performance for experienced pilots. In addition to their predictive validity, the following reasons should be considered for including an intelligence test in the set of tests used in a pilot selection system:

- Intelligence tests are relatively inexpensive and cost-efficient to administer. The traditional paper-and-pencil tests require no special apparatus and can be administered to large groups of applicants. Tests usually can be administered by an operator's human resources personnel. Computer-based intelligence tests may be more expensive but offer the advantage of immediate results;

- Many other tests used in pilot selection correlate moderately with scores on intelligence tests. In terms of overall predictive validity, intelligence tests often are the best predictors of pilot performance among all tests in a particular set; and,

- Many aircraft operators today find that the education level and background of pilot applicants is changing. For example, until relatively recently, many U.S. air carriers hired predominately ex-military pilots. These pilots previously had been selected through rigorous military processes,[17] and they had college degrees. U.S. air carriers no longer hire predominately ex-military pilots, and some do not require pilot applicants to have degrees from accredited colleges or universities. Air carriers in some other parts of the world have reported difficulties recruiting pilots at the ab initio level and the experienced level, and, like their U.S. counterparts, some have lowered their educational requirements and/or have begun recruiting from nontraditional sources.

  When educational requirements are lowered, however, some pilot applicants may not have the knowledge required to complete training within the required time. A carrier could expect to correct deficiencies by providing additional training. Nevertheless, some

individuals may not have the required intelligence to overcome the deficiencies in the allotted time. Intelligence testing is one of the few valid methods of identifying such individuals.

Motor Tests Best Suited to Ab Initio Pilot Selection

Tests that involve manipulation of some type of physical apparatus[18] are known as "motor tests" or "psychomotor tests." (They formerly were called "apparatus tests.") The distinction between a motor test and a psychomotor test is not standardized, but "psychomotor" generally implies that some type of cognitive processing is required to perform the test well. For example, a test that requires the applicant to tap his/her index finger as rapidly as possible would be called a motor test. A test that requires the applicant to place a pointer on a dot that moves at different speeds in a figure-eight pattern on a computer screen usually would be called a psychomotor test.

Motor tests typically are not used to select experienced pilots. Nevertheless, motor tests, which often use relatively inexpensive apparatus, may be useful for ab initio pilot selection in places where pilot applicants may have had limited exposure to technologically advanced machinery and computers. Psychomotor tests — which often measure eye-hand coordination — are included in test batteries for experienced pilots; for example, the set of tests used by Qantas contains a test of eye-hand coordination and a test of eye-hand-foot coordination.[19] Psychomotor tests also may be used in ab initio pilot selection batteries, such as one used during the 1990s by Cathay Pacific Airways, which contains an eye-hand coordination test and an eye-hand-foot coordination test.[20] Scores from the psychomotor tests used by Qantas correlated with flight-training performance reports that were obtained after the pilots were assigned to operational flying almost as well as scores from a simulator test that also was included in the set of pilot selection tests.

The major disadvantage of psychomotor tests is their expense. Additionally, although the Qantas results show a correlation between scores on the psychomotor tests and scores on the flight-training performance reports, they reflect data from only one carrier.

Aircraft operators should develop a structured pilot selection system carefully, should establish its predictive validity and then should determine whether adding a psychomotor test would be cost-efficient.

Information-processing Tests Measure Reaction Time

Information-processing tests measure the speed at which an individual performs various cognitive functions.[21] The distinguishing feature of these tests is their use of reaction time, measured in milliseconds, as the primary measure of performance. Because reaction time must be measured so accurately, these tests are conducted using computers.

For example, an information-processing test might be designed to determine how rapidly an individual can respond to a simple stimulus on a computer display. During the test, the numerals "1" through "8" would be shown one at a time, at random, on the display. The job applicant would be required to press the corresponding number on a keypad, and the computer software would measure how rapidly the test-taker responded to the appearance of each numeral. A total of 100 numerals might appear, and the 100 response times would be averaged to measure the test-taker's performance.

Many commercial software packages are available to conduct information-processing tests. Basic software packages include several different information-processing tests. More comprehensive software packages include separate tests and combinations of tests that are designed to assess timesharing skills.[22] In a timesharing test, the test-taker is asked to perform two information-processing tasks concurrently. The predictive validity of timesharing tests is slightly greater than the predictive validity of information-processing tests.

The most comprehensive software packages include psychomotor tests, combinations of information-processing tests, combinations of information-processing tests with psychomotor tests, and feedback (a response) about the individual's performance on the tests. The software can assess factors such as an individual's risk-taking traits and decision-making abilities, in addition to the speed of cognitive processing and timesharing. Some software may require specialized computers and peripheral equipment to administer the tests.

Information-processing tests have several characteristics that make them attractive for structured pilot selection systems. They have a "game-like" quality that pilot applicants typically enjoy. Because they are administered using a computer, the results are available immediately. When used under controlled conditions, scores on these tests are unaffected by gender and ethnicity.

The major disadvantage of these tests is their cost. Aircraft operators may be required to purchase proprietary hardware and software, and may be charged a fee for each applicant who takes the test.

Personality-trait Testing Involves Caveats

Tests to evaluate personality traits of pilots may be the least successful type of test in ethnically heterogeneous countries. For example, some reports on personality-trait testing in U.S.

military pilot selection systems say that such tests have not succeeded.[23,24]

"The use of self-report personality scores did not enhance the predictive validity of a selection system," said one report on personality-assessment tests used by the U.S. Air Force. "The analyses of personality variables under investigation by the Air Force show very little promise for use in selecting or classifying aviation candidates."

A study of the pilot selection system at Qantas, however, said that there were significant positive correlations between personality scales and performance scores in training and probation.[25]

European air carriers have had more success with tests to evaluate personality traits. A study of one European charter operator (the study did not identify the operator) measured an increase in the "classification accuracy" of its pilot selection system from 73.8 percent to 79.3 percent after a test to evaluate personality traits was added to the set of tests.[26] (Classification accuracy is the percentage of applicants whose performance on a pass/fail criterion was predicted correctly.) Two reports on SAS's pilot selection system — written more than 30 years apart — present data from personality tests. Both reports said that SAS used experienced psychologists to administer the tests and to interpret them.[27,28]

Tests to evaluate personality traits have been criticized for their culture-specificity. For example, data show a strong predictive validity for one test — the Defense Mechanism Test — in Scandinavia and an absence of predictive validity for the same test in the Netherlands and the United Kingdom.[29] Other data show that a test developed for Lufthansa Airlines in Germany had poor predictive validity in China to a pilot applicant's overall flight-training grade.[30] Researchers were unsure of the cause but planned to conduct further studies to compare tests developed in China with the one developed in Germany.

Some researchers also have criticized tests to evaluate personality traits for their "transparency" — that is, the ease with which job applicants determine the desired response to each question and then respond accordingly.[31,32,33,34] Their answers may reflect little about their real personalities; the predictive validity of the test is, therefore, questionable.

In the United States, legal factors in testing related to personality also must be considered. The Civil Rights Act of 1991 specifically prohibits differential scoring of items by ethnic group or gender. For example, if a question asks the job applicant about his/her favorite sports in secondary school and answers of "boxing," "diving" and "gymnastics" are known to correlate highly with success in training men, these answers could be given a higher score than other answers to this item. "Boxing" may have a negative correlation with success in pilot training for women, however. Nevertheless, under the civil rights law, this answer must receive the same score for both men and women. This restriction on differential scoring has limited the usefulness of personality tests for pilot selection in the United States.

Biographical inventories — types of questionnaires — sometimes are used instead of personality-related tests, particularly in the United States, to ask about the applicant's family, education, hobbies and sports interests.[35] Some questions also pertain to interest in aviation and career goals. Reponses to the items are correlated with criterion scores. Items that correlate positively with one or more of the criterion scores are included in the applicant's total score.

Interviews Remain Valuable Tools

Interviews may be the most common element of a pilot selection system used by aircraft operators. Interviews may be conducted by one person or by a board (group). The interviewers may be human resources specialists, psychologists or senior pilots. At large air carriers, a pilot applicant may have several individual interviews, as well as a board interview.

Studies have shown that structured interviews (in which the same questions are asked of all applicants and the responses are assessed) have greater predictive validity for job performance than unstructured interviews, perhaps because unstructured interviews typically do not present all applicants with the same questions and because they lack scoring guides.[36,37]

Cost is one problem associated with interviews conducted by one or more pilots who are on flight status. Some aircraft operators may underestimate the cost, as they pay not only the pilots' salaries and benefits but also lost-revenue costs.

Another problem is that interview questions may become known in advance to applicants, compromising spontaneous responses: Some aircraft operators' interview questions are available on the Internet for a fee, many books provide guidance on how to respond in an interview, and consultants may advise job applicants on the appropriate responses and appropriate demeanor for interviews at specific air carriers. The effect of this advance preparation on the predictive validity of interviews has not been established.

Flight-skills Tests Conducted on the Ground

Many aircraft operators require an assessment of pilot applicants' flight skills. Traditionally, these assessments were conducted in an aircraft, but today, they typically are conducted in a simulator or other flight-training device.

Some simulator assessments are based on national test standards for the appropriate level of pilot license (i.e., the airline transport license or an instrument rating), and scoring of the assessments sometimes presents problems. The predictive validity of a test is closely related to its reliability. If a test has a low reliability, its predictive validity also is low. For tests with scores based on observer ratings, reliability is measured in terms of "inter-rater reliability" (the correlation between the scores given by two check pilots rating the same pilot performance).

Two 1997 research reports said that inter-rater reliability typically is poor at air carriers. The reports said that reasons were not determined but that insufficient training of check pilots might be one cause.[38,39] Research on improving the inter-rater reliability of check pilots found that more intensive training of check pilots and periodic retraining would help improve inter-rater reliability.[40,41,42]

System Development Can Take Months

Aircraft operators typically make two mistakes in developing or revising a structured pilot selection system.

They do not allow sufficient time for the development process or revision process and, therefore, introduce the process too late in their hiring cycle. If an air carrier decides to develop a new pilot selection system, the process must begin before pilot hiring begins. Some air carriers, however, wait until a few weeks before hiring begins to consider changing the pilot selection system. Management should allow three months to six months to develop a new structured pilot selection system for a small-size air carrier to a medium-size air carrier. This allows sufficient time for system developers to schedule meetings with management, to make informed decisions about the criteria and tests, and to train testing personnel, check pilots and interviewers. A large air carrier that annually hires hundreds of new pilots may require considerably more time for the development process.

Revising an existing pilot selection system typically requires much less time than developing a new system. Three months is sufficient for making most changes.

The other common mistake is purchasing an expensive test without adequate consideration of the purpose of the test or the alternatives. Management should not purchase a test unless the test is demonstrably related to the KSAs or personality traits that an aircraft operator wants its pilots to possess. Even if a test is related to KSAs or personality traits of importance to the operator, adding the test may not increase the predictive validity of the selection system.

Management should not purchase a test solely because of the reported usefulness of the test to other aircraft operators; operators that appear similar may differ in subtle ways that affect the predictive validity of their tests. For example, although one air carrier may find that a particular test increases the predictive validity of its structured pilot selection system, another air carrier may find that the same test has no effect on the predictive validity. Aircraft operators who consider purchasing such tests should seek appropriate expertise to evaluate the vendor's claims and perform cost/benefit analyses.

For maximum effectiveness, tests, like other elements of the pilot selection system, should be developed to match an aircraft operator's specific requirements. In that way, the system will be able to identify the applicants most likely to succeed as pilots for a specific aircraft operator.♦

Notes

1. Guion, Robert M. *Assessment, Measurement and Prediction for Personnel Decisions.* Mahwah, New Jersey, U.S.: Lawrence Erlbaum Associates, 1998.

2. Hunter, David R.; Burke, Eugene F. *Handbook of Pilot Selection.* Aldershot, England: Avebury Aviation, 1995.

3. Trankell, Arne. "The Psychologist as an Instrument of Prediction." *Journal of Applied Psychology* Volume 43 (1959): 170–175.

4. Stead, Greg. "A Validation Study of the QANTAS Pilot Selection Process," chapter in *Human Resource Management in Aviation*, edited by Eric Farmer, Aldershot, England: Avebury Technical, 1991.

5. Gordon, Hans. "Selection of Ab Initio Pilot Candidates: The SAS System," chapter in *Human Resource Management in Aviation*, edited by Eric Farmer, Aldershot, England: Avebury Technical, 1991.

6. Gopher, Daniel. "A Selective Attention Test as a Predictor of Success in Flight Training." *Human Factors* Volume 24 (1982): 173–184.

7. Thorndike, Robert L. *Personnel Selection Test and Measurement Techniques.* New York, New York, U.S.: John Wiley and Sons, 1949.

8. Hunter and Burke.

9. Guion.

10. Cohen, Jacob. "The Cost of Dichotomization." *Applied Psychological Measurement* Volume 7 (1983): 249-253.

11. Schmitt, Neal; Chan, David. *Personnel Selection: A Theoretical Approach.* Thousand Oaks, California, U.S.: Sage, 1998.

12. Most current theories of intelligence say that human intelligence is composed of a number of abilities. (An intelligence test usually is composed of subtests that assess a number of abilities comprising human intelligence. The manager of a structured pilot selection system may choose to include an entire intelligence test or only a few of its subtests in the system. For example, a manager may decide to use only the "spatial relations" subtest of an intelligence test consisting of 10 subtests. Such subtests may be referred to as ability tests.)

    Another commonly misunderstood term is "stanine." This term refers to a method of scoring a test on a nine-point scale and is a contraction of the term "standard nine," [according to *Introduction to Classical and Modern Test Theory,* by Linda Crocker and James Algina. Fort Worth, Texas, U.S.: Harcourt Brace Jovanovich, 1986, and "U.S. Air Force Pilot Selection Tests: What Is Measured and What Is Predictive?" by Thomas Carretta and Malcolm J. Ree in *Aviation, Space and Environmental Medicine* Volume 67 (1996): 279–283]. During World War II, the pilot-assessment tests were scored using the stanine method. Pilots seem to have referred to the tests as "stanine tests," and gradually this term began to be used to refer to any paper-and-pencil ability test.

13. Guilford, J.P.; Lacey, John. *Printed Classification Tests Report No. 5.* Army Air Forces Aviation Psychology Program Research Reports. Washington, D.C., U.S.: U.S. Government Printing Office, 1947.

    Although this article examines civilian pilot selection systems for experienced pilots, national governments historically have funded much research on test development and on the identification and refinement of criterion measures to improve the military pilot selection system; many references in this paper reflect this evolution.

14. Stead.

15. Trankell.

16. Gordon.

17. Carretta, Thomas R.; Ree, Malcolm J. "U.S. Air Force Pilot Selection Tests: What Is Measured and What is Predictive?" *Aviation, Space and Environmental Medicine* Volume 67 (1996): 279–283.

18. Melton, Arthur W. *Apparatus Tests Report No. 4.* Army Air Forces Aviation Psychology Program Research Reports. Washington, D.C., U.S.: U.S. Government Printing Office, 1947.

19. Stead.

20. Bartram, Dave; Baxter, Peter. "Validation of the Cathay Pacific Airways Pilot Selection Program." *International Journal of Aviation Psychology* Volume 6 (1996): 149–170.

21. Wickens, Christopher D. *Engineering Psychology and Human Performance*, 2nd edition. New York, New York, U.S.: HarperCollins Publishers, 1992.

22. Damos, Diane L. "Using Meta Analysis to Examine the Predictive Validity of Single- and Multiple-task Measures to Flight Performance." *Human Factors* Volume 35 (1993): 615–628.

23. Dolgin, Daniel L.; Gibb, Gerald D. "Personality Assessment in Aviator Selection," chapter in *Aviation Psychology*, edited by Richard S. Jensen, Aldershot, England: Gower Technical, 1989.

24. Siem, Frederick M. "Predictive Validity of an Automated Personality Inventory for Air Force Pilot Selection. *International Journal of Aviation Psychology* Volume 2 (1992): 261–270.

25. Stead.

26. Hormann, Hans-Jurgen; Masche, Peter. "On the Relation Between Personality and Job Performance of Airline Pilots. *International Journal of Aviation Psychology* Volume 6 (1996): 171–178.

27. Trankell.

28. Gordon.

29. Martinussen, Monica; Torjussen, Tore. "Does DMT Predict Pilot Performance Only in Scandinavia?" In *The Proceedings of the Seventh International Symposium on Aviation Psychology*, edited by R.S. Jensen and D. Neumeister. Columbus, Ohio, U.S.: Ohio State University, 1993.

30. Hormann, Hans-Jurgen; Luo, Xiao Li. "Development and Validation of Selection Methods for Chinese Student Pilots." In *Tenth International Symposium on Aviation Psychology*, edited by Richard Jensen, et al. Columbus, Ohio, U.S.: Ohio State University, 1991.

31. Hunter and Burke.

32. Dolgin and Gibb.

33. Siem.

34. Anastasi, Anne; Urbina, Susana. *Psychological Testing,* 7th edition. Upper Saddle River, New Jersey, U.S.: Prentice Hall, 1997.

35. Hunter and Burke.

36. McDaniel, Michael A., et al. "The Validity of Employment Interviews: A Comprehensive Review and Meta-analysis." *Journal of Applied Psychology* Volume 79 (1994): 599–616.

37. Janz, Thomas; Hellervik, Lowell; Gilmore, David C. *Behavior Description Interviewing.* Upper Saddle River, New Jersey, U.S.: Prentice-Hall, 1986.

38. Baker, David P.; Dismukes, R. Key. "Tools for Investigating Rater Reliability and Accuracy." In *Proceedings of the Ninth International Symposium on Aviation Psychology*, edited by Richard S. Jensen. Columbus, Ohio, U.S.: Ohio State University, 1997.

39. Williams, Donna M.; Holt, Robert W.; Boehm-Davis, Deborah A. "Training for Inter-rater Reliability: Baselines and Benchmarks." In *The Proceedings of the Ninth International Symposium on Aviation Psychology*, edited by Richard S. Jensen. Columbus, Ohio, U.S.: Ohio State University, 1997.

40. Baker and Dismukes.

41. Williams et al.

42. Hamman, William R.; Beaubien, J. Matthew; Holt, Robert W. "Evaluating Instructor/Evaluator Inter-rater Reliability From Performance Database Information." In *The Proceedings of the Tenth International Symposium on Aviation Psychology*, edited by Richard Jensen, et al. Columbus, Ohio, U.S.: Ohio State University, 1999.

About the Author

Diane L. Damos, Ph.D., has worked in the area of pilot selection for more than 30 years. She received a doctorate in aviation psychology from the University of Illinois in 1977 and taught from 1977 until 1997 at the State University of New York, Arizona State University and the University of Southern California. She has lectured and has taught courses on pilot selection in Taiwan, South Africa, Spain and Canada. In 1995, she founded Damos Aviation Services.

Damos has conducted research on pilot selection for the U.S. Air Force and the U.S. Navy and helped develop the computer-based testing used by the Air Force. She has consulted with national governments and air carriers on military pilot selection and civilian pilot selection. Damos has received numerous grants and contracts from the Navy, Air Force and the U.S. National Aeronautics and Space Administration to support her research. She has written more than 25 articles in scientific publications and 50 technical reports and proceedings articles on training, cognition and pilot selection systems. She is a consulting editor for the International Journal of Aviation Psychology *and a fellow of the Human Factors and Ergonomics Society. She served on the National Academy of Sciences/National Research Council Panel of Human Factors in Air Traffic Control Automation. She also has been an expert witness in many aviation-related court cases on pilot certification, training and performance.*

Damos holds a commercial pilot certificate with ratings in single-engine airplanes and multi-engine airplanes and an instrument rating, an advanced ground instructor license and an instrument ground instructor license.

Further Reading From FSF Publications

Global Aviation Information Network (GAIN) Aviation Operator Safety Practices Working Group. "Operator's Flight Safety Handbook." *Flight Safety Digest* Volume 21 (May–June 2002).

Shappell, Scott A.; Wiegmann, Douglas A. "Human Factors Analysis and Classification System." *Flight Safety Digest* Volume 20 (February 2001).

Sallee, G.P.; Gibbons, D.M. "Propulsion System Malfunction Plus Inappropriate Crew Response (PSM+ICR)." *Flight Safety Digest* Volume 18 (November–December 1999).

Weiner, Earl L.; Chute, Rebecca D.; Moses, John H. "Transition to Glass: Pilot Training for High-technology Transport Aircraft." *Flight Safety Digest* Volume 18 (June–August 1999).

Pinet, Jean; Enders, John H. "Human Factors in Aviation: A Consolidated Approach." *Flight Safety Digest* Volume 18 (May 1999).

FSF Icarus Committee. "Aviation Safety: Airline Management Self-audit." *Flight Safety Digest* Volume 18 (May 1999).

National Civil Aviation Review Commission. "A Safe Flight Into the Next Millennium." *Flight Safety Digest* Volume 17 (January 1998).

Mears, Allen. "Education — The Corporate Possibilities." *Flight Safety Digest* Volume 9 (January 1990).

"The Future of the Civil Aviation System." *Flight Safety Digest* Volume 9 (January 1990).