Federal Aviation Administration

DOT/FAA/AM-19/6
Office of Aerospace Medicine
Washington, DC 20591

Best Practices in Pilot Selection

Dana Broach

David Schroeder

Kevin Gildea

February 13, 2019

Final Report

NOTICE

____

This publication and all Office of Aerospace Medicine technical reports are available in full-text from the

Civil Aerospace Medical Institute's publications website: http://www.faa.gov/go/oamtechreports

Technical Documentation Page

1. Report No. DOT/FAA/AM-19/6	2. Government Accession No.	3. Recipient's Catalog No.
4. Title and Subtitle Best Practices in Pilot Selection		5. Report Date February 13, 2019
		6. Performing Organization Code
7. Author(s) Dana Broach, Kevin Gildea, & David J. Schroeder		8. Performing Organization Report No.
9. Performing Organization Name and Address Aerospace Human Factors Research Laboratory AAM-500 P.O. Box 25082 Oklahoma City, OK 73125		10. Work Unit No. (TRAIS)
		11. Contract or Grant No.
12. Sponsoring Agency Name and Address Office of Aerospace Medicine Federal Aviation Administration 800 Independence Avenue, S. W. Washington, DC 20591		13. Type of Report and Period Covered Final
15. Supplementary Notes		

16. Abstract

Demand for new civilian pilots continues to grow as the world's aviation system expands. U.S. mainline and regional air carriers will need about 1,900 to 4,500 new pilots per year (U.S. Government Accountability Office, 2014). Selection of the new pilots is a critical human resource management challenge for operators. We review civilian pilot selection procedures relative to a set of seven best practices: 1) Conduct a job analysis; 2) Define measurable, observable job performance metrics; 3) Identify and use reliable and valid predictors; 4) Conduct an appropriate validation study; 5) Determine cut-scores (pass/fail) on tests based on predicted job performance; 6) Evaluate the fairness of tests and cut-scores; and 7) Document the analyses. We reviewed 15 pilot test batteries and selection processes for 2 U.S. and multiple foreign airlines based on publicly available information. Overall, adherence to best practices as defined by the relevant legal and professional guidelines, standards, principles, and practices for the development, validation, and use of employee selection procedures, was inconsistent. Relatively few of the test batteries were based on a recent or current job analysis. There is evidence of psychometric reliability and useful validity for the test batteries. However, there is no evidence on the reliability or validity of the interviews used in pilot selection. The primary job performance criterion is success or failure in training. No data are available on the comparative selection (pass) rates by demographic groups on the test batteries. Technical reporting and documentation on pilot test batteries is better in Europe than in the U.S. Overall, we agree with the 2012 conclusion reached by the International Air Transport Association (IATA) that pilot selection programs worldwide and in the U.S. do not appear to have a strong scientific basis. We make two recommendations to improve the state-of-the-art in civilian pilot selection.

17. Key Words	18. Distribution Statement
Pilot Personnel Selection Job Performance Job Analysis Ability Tests	Document is available to the public through the National Technical Information Service, Springfield, VA 22161

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price
UNCLASSIFIED	UNCLASSIFIED	73	

Form DOT F 1700.7 (8-72) Reproduction of completed page authorized

ACKNOWLEDGMENT

Research report in this paper was conducted under the Air Traffic Program Directive/Level of Effort Agreement between the Human Factors Division (ANG-C1), FAA Headquarters, and the Aerospace Human Factors Research Division of the FAA Civil Aerospace Medical Institute.

TABLE OF CONTENTS

TABLE OF FIGURES

TABLE OF TABLES

BEST PRACTICES IN PILOT SELECTION

INTRODUCTION

Pilot selection has been a topic of intense study for about 100 years in both the military and civilian worlds. Much of the research has focused on military pilot selection. But civilian pilot selection has become more important in the last 20 years with the growth of air transportation around the globe. Many new civilian pilots will be required over the next 20 years. For example, Boeing (2014) projects 633,000 new pilots worldwide will be needed in commercial aviation over the next 20 years. In North America, Boeing projects a need for 88,000 new pilots by 2034. According to the United States Government Accountability Office (GAO, 2014), there are about 72,000 air transport pilots flying for 15 scheduled mainline airlines and 70 regional airlines in the United States. Based on a review of various workforce studies in aviation, the GAO (p. 16) projected that 1,900 to 4,500 new pilots would be needed, on average, each year over the next ten years, to replace pilots who retire or otherwise leave the cockpit. Stated in a different way, U.S. mainline and regional operators will need to hire 19,000 to 45,000 new pilots over the next ten years. Selecting the "right" pilot from among many applicants to fill behind expected losses and to accommodate growth will be a significant human resource management challenge for both mainline and regional carriers for many years to come (Duggar, Smith, & Harrison, 2011).

A critical element in meeting that challenge will be an adequate supply of qualified applicants from which to make selections into the cockpit (Croft, 2015a, b). Studies of the pilot labor market suggest four factors will significantly influence the size of the pool of qualified pilots from which airlines can select new pilots to replace retirees and other losses: 1) Industry growth, 2) pay, 3) education costs, and 4) Federal Aviation Administration (FAA) certificate requirements (Smith, Herchko, Bjerke, Niemczyk, Nullmeyer, Paasch, & NewMeyer, 2013; Higgins, Lovelace, Bjerke, Lounsberry, Lutte, Friedenzohn, …Craig, 2013; Lake, 2011; Smith, Bjerke, NewMeyer, Niemczyk, & Hamilton, 2010; Moak, 2014).

The first factor influencing the supply of pilots is industry growth. The aviation industry is growing worldwide, fueling demand for new pilots. Passenger enplanements are projected to increase by 2.2% over the next 20 years (FAA, 2016). However, airlines are also aggressively managing capacity (available seats and flights) to maximize revenue. For example, departures (a proxy for number of flights conducted) are expected to increase by just 1.3% over the 20-year FAA forecast horizon (FAA). To accommodate growth in the number of passengers, carriers will likely fill a larger proportion of seats in somewhat larger aircraft (for example, 70- rather than 50-seat regional jets) without adding flights. Another factor is continuing consolidation of operators through mergers. For example, several airlines have merged in the last ten years to create three large network or mainline carriers: Delta with Northwest in 2008; United with Continental in 2010; and American with U.S. Airways in 2013. These three large carriers and Southwest Airlines account for 76% of the U.S. domestic market, based on revenue passenger miles (FAA, 2016, p. 9). One stated goal for mergers is optimization of fleet size relative to demand, which has the net effect of constraining the actual number of airplanes flying, and thus, the number of pilots required by the industry (Schnurman, 2015). The large carriers in the U.S. are also exploring ways to increase available seats without increasing the number of flights. For example, some airlines have reduced the average seat pitch (distance between seats), particularly in the "economy" cabin. Other airlines have installed new, thin seats that add capacity without requiring additional flights. So while the number of passengers might increase over the next decade (about 2%) as forecast by the FAA, the number of flights carrying those passengers will grow at a slower rate (about 1%), thus potentially limiting to some degree the number of new pilots required to replace retirees and accommodate airline growth.

The second and perhaps controversial factor in determining the size of the future pilot labor pool is pay. The Air Line Pilot's Association, International (ALPA, 2015, 2016) specifically points to pay as a major factor in the apparent pilot shortage, particularly for the regional airlines. The GAO presented a more nuanced view in 2014, reporting that available data indicated that a large pool of qualified pilots (e.g., those with an ATP certificate and appropriate medical certificate) existed relative to anticipated demand for new pilots (p. 11). The GAO noted, however, that it was unknown whether such pilots were willing or available to work at the wages being offered by the regional airlines. On one hand, pay at senior levels in the airlines is very attractive. The median pay for airline pilots, according to the Bureau of Labor Statistics, was $114,200, $154,100 at the 75th percentile, and $187,900 at the 90th percentile as of 2014 (most recent data available; U.S. Bureau of Labor Statistics, 2015). On the other hand, pay at the entry level with regional airlines is much less attractive, especially with consideration of the costs of obtaining the necessary ATP certificate.

The costs of obtaining the pilot certificates, including flying time, required by the FAA to work for an airline, can be substantial and acts as the third constraint on the pilot supply pipeline. For example, one route is to earn a bachelor's degree at a school with an accredited pilot training program. At a private university, the cost of the bachelor's degree can approach or exceed $100,000, with additional costs for flight instruction (GAO, 2014). Educational costs are often substantially less at U.S. public universities with pilot training programs but still run in the tens of thousands of dollars. Other estimates run as high as $50,000 per year cost for a four-year college education with flight training (Croft, 2015a). Other paths to the ATP certificate, such as non-collegiate flight schools, have significant costs to the would-be pilot. These costs, in turn, are perceived as dampening enrollments in pilot training programs despite the long-term earnings potential (Croft, 2015b).

The final factor that will shape the supply of pilots is certification. The ATP certificate is a legal requirement. The simple fact is that the U.S. Congress passed a law mandating that pilots of aircraft operating under the regulations governing air carriers (Part 121 of title 14, Code of Federal Regulations (14 C.F.R. § 121)) *must* hold an ATP certificate (Public Law 111-216 § 216, August 1, 2010, 124 Stat. 2367). The FAA was bound by the statute (law) to issue a rule (regulation) to that effect no later than August 1, 2013. FAA published the final rule on July 15, 2013 (*Federal Register*, *78*(135), 42324-42380). The impact of Section 216 of Public Law 111-216 on the supply of pilots has been debated. For example, Lake (2011) concluded, prospectively, that the law and subsequent rule would have a detrimental effect on the pilot supply. The Regional Airline Association (RAA) described the rule's impact as adding extraordinary "time and financial burden" to pilots seeking to enter the workforce and as resulting in the loss of service in smaller markets due to pilot unavailability (RAA, 2014). The National Association of Flight Instructors (2012, p. 2) asserted that "[I]n light of new rulemaking, it is unlikely that the flight training industry will be able to offer a continuous supply of qualified pilot to meet the demands of commercial carriers" (Blair & Freye, 2012). Other analyses are less pessimistic. For example, McGee (2015) developed a comprehensive ATP supply and demand model to assess future requirements. He concluded that the rule would “only minimally affect flow capacity of the [pilot] pipeline” (p. 81). McGee specifically modeled flow through the “Part 61” and “Part 141” flight schools (the distinctions are discussed later in this report). He concluded that while the short-term effect was to stretch out times in flight schools, there will be “enough ATP certified pilots to fulfill the demand at the majors and minors for the next ten years” (p. 78). McGee also found that the long-term effect of the rule would be a shift away from Part 61 flight schools toward Part 141 programs associated with colleges and universities (p. 81).

Even with constrained supply relative to demand for pilots, it will still be important to use sound, scientifically-based processes for their selection, as there are significant costs associated with wrong decisions. One obvious cost of improper selection is the risk of catastrophe. For example, failures in basic airmanship such as understanding how to react to an impending or actual aerodynamic stall, have been implicated in accidents such as Colgan 3407 and Air France 447. Selecting pilots with better airmanship can increase the margin safety; conversely, failing to eliminate pilots with poor airmanship can increase risks to safety and property. Such hazards to safety and property elevate the importance of sound selection procedures. Another obvious cost is unnecessary attrition from training required by the mainline or regional carrier. Those dollars invested in training a pilot who then fails are essentially avoidable costs to the employer. A scientifically-based pilot selection process might reduce such avoidable costs.

Moreover, as Damos noted in 2014, pilot selection has been complicated by two factors over the last two decades. First, despite over 100 years of research on pilot selection, the dissemination and application of research findings to pilot selection around the world is a continuing problem. Part of the issue is in communicating psychological and statistical concepts, data, and recommendations in a language that corporate managers can understand, believe, and use. Second, at least in the U.S. market and probably in other non-U.S. employment markets, there has been a pronounced shift in the sourcing of pilots (Damos). Following World War II and into the early 1980s, former military pilots were a dominant source for pilots. U.S. Military pilots were the product of a rigorous, intensive, and unrelenting winnowing process. Those that left the military for civilian seats in civilian airliners had survived that winnowing. In selection terms, they had been pre-selected on job-relevant factors such as cognitive ability, emotional stability, and demonstrated airmanship. Now, former military pilots account for a minority of new pilots in the U.S. More of the pilots hired in the last 20 years did not experience the level of scrutiny given to persons aspiring to be a military pilot. Indeed, it could be argued that entry into non-military flight training programs today is more dependent on the “wallet test” than any other single factor. As a result, the variability between aspiring civilian pilots in terms of factors such as cognitive abilities, attitudes, personality, and airmanship is likely to be much greater than in the past. That wider range of variability reinforces the need for improved pilot selection processes to ensure hiring of pilots from the upper rather than lower ends of the knowledge, skill, and ability distributions.

With these changes in the industry and characteristics of pilot applicants comes renewed interest in evaluating selection procedures for those interested in pursuing the ATP certificate and flying in revenue service. Prior research has explored in detail many of the critical characteristics of the selection tools currently in use. However, as with much research in applied settings mapping the selection instruments to the criteria found in the operational environment is challenging and complex. Much of the research in the civilian sector contains information that is necessarily proprietary and thus, is difficult to obtain for cross-program analysis. It is also difficult to follow and measure pilot performance as a career progresses from selection, through flight school, and across the pilot's career. Thus, many questions remain in terms of the effectiveness of civilian pilot selection procedures including:

- To what degree do selection measures employed accurately predict performance in training and in daily flight operations?
- What are the safety, organizational, fiscal, and operational risks associated with errors in pilot selection?
- In current training programs, what additional selection tools or performance metrics are employed?
- What are the near optimal places in training for additional testing, what should those tests measure, and what, if any, remedial training is viable?
- How do the selection procedures vary between air carrier, university, and Fixed Base Operator (FBO) flight schools under Parts 61, 141 and 142?

The focus of this report will be the identification of "best practices" through a review of literature addressing selection systems in the field of aviation specifically and transportation in general.

Organization of report

The report is organized into five chapters. In the first chapter, we review the technical and legal framework for personnel selection (screening) programs in the United States. This framework has evolved through 40-plus years on interplay between applied psychological research and the developments in equal employment opportunity jurisprudence. The review in the first chapter is used to develop a list of "best practices" against which pilot and other occupational selection programs can be evaluated. The second chapter describes the major pathways, in the US, to becoming a pilot. The third chapter begins with a brief summary of previous reviews of civilian pilot selection. Then several pilot selection (screening) batteries are described, based on publicly available information. These pilot-specific selection programs are evaluated with respect to the "best practices" described in the first chapter. This third chapter also includes a discussion of testing for crew resource management (CRM). The fourth chapter turns to examples of other transportation-related occupational selection programs such as air traffic controllers, again based on publicly available documentation. In the fifth chapter, we summarize our evaluations relative to the "best practices" framework, make recommendations, and pose additional research questions.

Terminology

"Screening," "selection," "applicant," "candidate," "recruitment," and "hiring" are words often used in discussions of the processes for attracting, evaluating, and hiring new persons into an organization, generally without any specific definition or understanding. To avoid misunderstandings, we will use key words such as "selection" and "applicant" with very specific operational definitions throughout this report. "Recruitment" for this report refers to the organizational processes for advertising that vacancies exist within an organization in a given occupation. Recruitment informs potential applicants of the nature of work performed in that occupation. Recruitment also provides information about factors such as pay and benefits, locations/duty stations, and perhaps qualification requirements such as certificates, licenses, and experience. Recruitment, in this report, is the employer activity of providing or giving information about a job to persons who might be interested. Recruitment in this usage does *not include evaluation* of any individual vis-à-vis the requirements of the position to be filled. "Applicant" for purposes of this report refers to a person who has expressed an interest in being considered for an open position (a vacancy) with an employer, as evidenced by the submission of an actual application. As with recruitment, the term "applicant" does not include any evaluative component or aspect. A person is an applicant until (a) he or she is excluded or removed from the hiring process for any reason at any stage or (b) a *bona fide* job offer is extended by the employer. Once a *bona fide* job offer is extended, we will refer to that person as a candidate for employment. If the candidate clears all subsequent hurdles, is extended a *bona fide* final job offer, accepts the offer, reports for duty with the employer, and begins receiving pay and benefits from that employer, then that person will be referred to as an "employee" within the scope of U.S. law.

"Selection" and "screening" are functionally equivalent, in that an applicant is assessed *vis-à-vis* job requirements. Therefore, the broader term "selection" will be used throughout the report to encompass every step in the employer's assessment of an applicant, where such assessments result in the elimination of one or more applicants and the progression of other applicants in the selection process. For example, applications for a pilot's position with an air carrier might be reviewed initially for possession of an Air Transport Pilot (ATP) certificate and a valid medical certificate, both required for a pilot to operate an aircraft for a domestic, flag, or supplemental air carrier under U.S. federal regulations (see 14 C.F.R. § 61.23). Applicants without the ATP and valid medical certificates would be eliminated from further consideration; this decision ("Does the applicant have the required credentials?") is, in fact, a personnel selection decision within the ambit of the federal *Uniform Guidelines on Employee Selection Procedures* (29 Code of Federal Regulations (C.F.R.) § 1607) and the *Civil Rights Act of 1964* as amended (14 United States Code (U.S.C.) § 2000, *et seq*.). This credentials check is often referred to as "screening" applications (or applicants). But functionally, it is a personnel selection decision, in that some applicants proceed in the process and others do not.

The next step in the overall selection process might be to evaluate applicant cognitive abilities and personality with a computerized test battery. Again, some applicants might be eliminated from further consideration on the basis of their test scores on the test battery and other applicants might continue in the selection process. This represents another selection decision made by the employer. The remaining applicants might then be interviewed by the air carrier's Chief Pilot as the next step in the selection process. Each decision in the process that results in the elimination of one or more applicants is a selection decision. Therefore, we will use the term "selection" throughout the report to refer to these *employer* decisions.

As noted previously, an applicant who passes all hurdles in the selection process and is extended a *bona fide* tentative job offer will be labeled a "candidate" for employment. A candidate does not become an employee, that is, is not "hired," until (a) a *bona fide* final job offer is extended to the candidate, (b) the candidate accepts the final offer of employment, and (c) the person enters on duty with the employer in pay status.

There often is some confusion around academic, employer-sponsored, and employer-paid training programs. Terms like "student" and "trainee" enter into the discussion. The distinctions between academic, employer-sponsored, and employer-paid training are irrelevant to recruitment, as recruitment is about providing information. Similarly, in discussing academic training programs (such as universities and private vocational (flight) training schools for which the individual pays), applicants are persons who submit an application to enter into the training program. Decisions made by training institutions about which applicants to accept into training and which to exclude from training are *de facto* selection decisions, but are called “admissions.” Admissions into a training program are functionally equivalent to selection decisions, in that some persons progress into the training program and others do not. Persons who receive a *bona fide* acceptance from a training institution are candidates. Persons who accept a *bona fide* offer of acceptance into a pilot training program *and* report for training will be labeled “students” rather than employees. The key difference is that students do not receive pay from an airline employer while enrolled in the pilot training program. So, for example, if a person applies to a university-based academic pilot training program, is accepted, and enrolls in the school but is not paid by an air carrier or other certificated operator, the person will be referred to as a student. In contrast, a person who is admitted into a pilot training program conducted or sponsored by a certificated air carrier *and* receives pay from that operator while in the pilot training program, then that person will be referred to as an “employee” of that air carrier. Where appropriate, such an employee will be referred to as an employee in training, not as a student. The key test is whether the person receives a W-2 or Form 1099 for taxable compensation from a certificated air carrier while enrolled in the training program. This is important in the U.S. as different laws and regulations apply to students and employees. Specifically, the *Uniform Guidelines on Employee Selection Procedures* and Title VII of the Civil Rights Act (14 U.S.C. § 2000 *et seq.*) apply to employers and employees but do not apply to educational institutions and students. Equal opportunity in education is governed by Title IX of the Civil Rights Act (codified at 20 U.S.C. § 1681 *et seq.*).

Finally, the overall purpose of this paper is to develop a set of “best practices” for pilot selection and compare pilot selection programs against those best practices. We have framed “best practices” for purposes of this review very specifically and intentionally in terms of compliance with U.S. legal constraints on personnel selection practices. These constraints impose a degree of scientific, professional, and procedural rigor on the development, validation, use, and evaluation of pilot selection programs. Pilot selection processes developed within this best practices framework are more likely to survive external (judicial) review and to produce superior organizational results than selection programs that are *ad hoc*, casual, ill-structured, poorly documented, and infrequently (if at all) evaluated with respect to costs and benefits to the employer.

CHAPTER 1: BEST PRACTICES IN PERSONNEL SELECTION PROCESSES

A strong technical framework for personnel selection has evolved in the United States over the past 40 years through the interplay of applied psychological research and equal employment opportunity jurisprudence. The framework applies to both private sector employers with 15 or more employees and to federal, state, and municipal agencies. The framework for "best practices" in personnel selection programs in the United States is defined by five elements. The first is the 1978 federal *Uniform Guidelines on Employee Selection Procedures* (29 C.F.R. § 1607). The second element is the 2014 professional *Standards for Educational and Psychological Testing* authored jointly by the American Educational Research Association (AERA), the American Psychological Association (APA), and the National Council on Measurement in Education (NCME). The third element is the professional *Principles for the Validation and Use of Employee Selection Procedures* published by the Society for Industrial and Organizational Psychology (SIOP, 2003). Professional practices and research is the fourth element of the technical framework. Finally, employment discrimination (especially the *Civil Rights Act of 1964,* as amended, 42 U.S.C. § 2000 *et seq.*) and case law are the fifth element of the technical framework for the development, validation, and evaluation of employee selection programs. This well established technical framework defines "best practices" in personnel selection in the U.S.

Conduct a job analysis

Given a decision to implement a new or changed personnel selection process, the first step is to conduct a job analysis. Job analysis is nothing more (or less) than a systematic approach to (a) *describing* the work done (or to be done) by a person in the target occupation or job, (b) *prescribing* the standards to which that work must be done, and (c) *ascribing* the attributes (e.g., the knowledge, skills, abilities, and other personal characteristics (KSAOs) or competencies) required to perform the work to the prescribed standard(s).

There is no single technical approach to a job analysis. Techniques can range from a simple review of existing job documentation to a labor-intensive detailed task analysis resulting in a flight procedures manual describing the how, when, and why of every interaction of the pilot with the Flight Management System (FMS) under almost every conceivable scenario likely to be encountered in flying the aircraft in revenue service. However, job analysis in support of personnel selection for jobs with public safety responsibilities is strongly influenced by case law in the United States. There are four major steps in job analyses that have survived judicial reviews in equal employment opportunity litigation (Guttman & Dunleavy, 2012; Malos, 2005; Thompson & Thompson, 1982). The first step is to review available information on the target occupation or job. Available information might include current position/job descriptions, previous job analyses, training curriculum materials, performance standards for the positions or jobs, aircraft- and company-specific operating procedures and training materials, and any relevant national standards for licenses or certificates. The goal of the review of available job information is to derive catalogs (lists) of job duties and/or tasks and KSAOs. The second step is to translate those derived work and KSAO statement into a job analysis survey to be completed by current job or position occupants (incumbents) and, often, supervisors and managers with responsibility for those jobs or positions. Typically, the work statements are rated on dimensions such as difficulty, importance, and frequency of performance. KSAO statements are often rated on dimensions such as the level of mastery required, the importance of the KSAO to successful performance, and when the KSAO is needed (at the time of hire or after, for example). The ratings data are analyzed to identify (a) the most important or critical duties or tasks and (b) the KSAOs that are most important to overall successful performance of the job and are also needed at the time of hire. The third step is to link specific KSAOs to successful performance of specific work duties or tasks. The analytic question is "To what degree is this specific KSAO essential to the performance of this specific duty or task?" Results of the linkage exercise provide a blueprint for the attributes to be assessed in the personnel selection program. The fourth step is to document the job analysis (Guttman & Dunleavy, 2012; Thompson & Thompson, 1982).

A key point in the job analysis is to consider the broader organizational context in which pilots operate in relation to co-workers, passengers and other customers, and external parties such as air traffic controllers. For example, pilots might be required by an airline to greet passengers as they board, suggesting a need for a customer service-oriented competency. Similarly, crewing practices for an airline might drive emphases on particular CRM approaches, suggesting a need to consider interpersonal capabilities in the job analysis.

Best practice #1: Conduct and document an analysis of the target job or occupation through systematic review of available information, development, and review of work and competency (KSAO) requirements, and explicit linkage of KSAOs to specific critical and/or important job duties and tasks. The analysis should take into account the organizational context and associated competency requirements of the specific organization. That is, each organization should conduct ***their own*** *job analysis.*

Define job performance measures

Given a job analysis, the next step is defining job performance criteria. One of the classic "traps" in the development and validation of employee selection procedures is to simply use whatever data are available on performance. Classic examples of available job performance measures include tenure, promotions, training completions, sales, attendance, and annual performance reviews. These measures are often readily available and organizationally important. An alternate approach (rarely taken) is to consider each major job duty and ask a simple question: How does the employer know that a pilot has performed this specific duty properly, that is, to what standard? The result is a catalog of potential job performance factors that, taken together, enable the employer to differentiate the good and the best pilots. The catalog of job performance factors can then be prioritized by factors such as impact on safety, avoided and incurred costs, and possibly impact on revenues and customer satisfaction. The highest value factors can be taken as candidate criteria and further evaluated for feasibility. Some high-value factors might not be feasible due to financial, operational, and organizational constraints. But the survivors of this winnowing process are more likely to represent the dimensions of performance most highly valued by the employer. The last step is to determine how to measure those high-value factors. Again, some performance factors might drop out because actually measuring a particular factor might not be feasible due to technological or other considerations. Yet the systematic inquiry into what pilot behaviors are valued by the employer and how to measure those is likely to produce meaningful, useful, and interpretable job performance criteria on which pilots can be measured and ranked.

Best practice #2: For each major job duty or activity, define measurable, observable performance metrics and set thresholds for acceptable performance on the metrics associated with those major job duties or activities.

Identify predictors

Once job performance criteria have been identified, attention turns to identification of predictors, that is, measures of the essential KSAOs. These predictors can take many forms, from written tests of, for example, knowledge of aviation regulations and flight procedures to scenario-based assessment in a flight simulation training device (FSTD). The predictors might be commercially available, "off the shelf," computerized tests or in-house, custom-developed assessments. Measures can also be modified or adapted to the requirements of a particular organization. Regardless of form or source, the predictors must be reliable measures of one or more essential KSAO. Reliability is a technical concept defined in the *Standards for Educational and Psychological Testing* (American Educational Research Association, *et al.*, 2014). Reliability can refer to (a) the degree to which scores on an instrument at time 1 and time 2 are similar (test-retest reliability or temporal stability), (b) the degree to which scores on version 1 and version 2 of an instrument are similar (parallel form reliability), and (c) the degree to which the items comprising the instrument are homogenous and consistent with one another (internal consistency). In the case of instruments based on observations and scoring of behaviors, reliability can also refer to the degree to which the observers agree in their observations (see Guion, 1998). Reliability does *not* encompass the relationships of scores on a given measure of an essential KSAO to measures of job performance.

Best practice #3: Identify, adapt, or create reliable measures of the KSAOs essential to successful performance of critical and/or important job duties or activities.

Evaluate relationship of KSAO measures to job performance criteria

The next step in development and validation of an employee selection procedure within this technical and legal framework is to evaluate the relationship of the predictors (measures of essential KSAOs) with the job performance criteria. Generally speaking, this involves collecting data for both predictors and job performance criteria and an analysis to estimate the magnitude of the relationship, often expressed as a correlation coefficient ("validity coefficient"). This process is known as criterion- or construct-related validation. Criterion-related validation refers to analysis of the relationship of one or more predictor scores to job performance measures. In contrast, in construct-related validation, the focus is first on the predictor as a measure of some latent (unobservable) attribute of a person (a "construct"), and then second on the relationship of that predictor score (as a measure of an underlying trait or characteristic) to the job performance measure. Another aspect of the validation strategy is when and from whom the data are collected. In a concurrent criterion-or construct-related validation study, the proposed predictors are administered to current job incumbents and job performance data are collected concurrently. In a predictive criterion- or construct-related validation study, the predictors are usually administered to job applicants but not used in the actual selection (hiring) decision. Then, at some future point in time, job performance data are collected for those that were hired, and the statistical relationship between predictors and criteria is investigated.

There are alternative approaches to validation (see McPhail, 2007). For example, the content of proposed predictors can be directly compared to critical or important work behaviors in what is known as content-oriented validation. Here, the exercise is rational rather than empirical. The classic example is a word processing (typing) test for an office automation specialist (e.g., secretary), where the job is preparing documents using the word processing application on a personal computer. Other approaches such as validity generalization might be used to argue for the validity of a proposed selection battery.

However, the goal of these different technical approaches is the same: Demonstrate the validity of the proposed predictors. Validity, under the *Standards for Educational and Psychological Testing*, "…refers to the degree to which evidence and theory support the interpretations of test scores entailed by the proposed uses of tests" (p. 9). In personnel selection, validity refers to the degree to which a meaningful inference about the level of future job performance can be made on the basis of predictor score(s). That is, do persons with higher scores on a predictor also perform better at work? Do persons with lower scores on a predictor perform more poorly at work? Validation provides answers to these questions.

Best practice #4: Evaluate the relationship of the measures of essential KSAOs with the job performance criteria. This is typically accomplished through either a predictive or concurrent criterion- or construct-related validation study.

Set cut-scores and/or decision rules

Given scores on the reliable assessments of KSAOs essential to the performance of critical and/or important job duties and tasks and evidence of the validity of those assessments, the next step in the development of a pilot selection process is to determine how those scores will be used to make determinations about which applicants proceed in the selection process and which are winnowed out. There are three decisions the employer must make in a personnel selection process. The first decision is the pass/fail (accept/reject) score for each assessment or test. There are many technical approaches to setting cut-scores (see Kehoe & Olson, 2005 for a review). The second decision is if scores from multiple assessments are to be combined into a single composite score, and if so, how those scores will be combined (see, for example, De Corte, Lievens, & Sackett, 2007). The third decision is how to use any composite score (or multiple scores if not combined) in making decisions about individual applicants. There are technical considerations in each decision about how to use the scores in making decisions about individuals. These technical considerations include the size of the standard errors of measurement for each score (e.g., the precision of the score), test (score) reliability and validity, and what trade-offs the employer is willing to accept. Kehoe and Olson (2005) provide a comprehensive review of these technical issues and relevant litigation in the U.S. Demonstrating a rational link between cut scores (whether on individual tests or a composite) and job performance is essential, where the lowest acceptable score is commensurate with lowest acceptable job performance. If a composite score is created, there should also be some rational nexus between the weights given to specific scores (to compute the composite) and the importance of the KSAO underlying those scores. Finally, the use of the cut-scores should be consistent with documentation of the rationale for any changes over time.

Best practice #5: Determine cut-score (pass/fail) on tests based on expected job performance at that score. Provide guidance in the use of the cut-score in making decisions, and document the rationale for changes in cut-scores over time.

Evaluate the fairness of the selection process

“Fairness” in personnel selection in the United States refers generally to the ratios of the selection rates for different groups. The selection rate is the proportion of persons “passing” a given hurdle in a selection system. For examples, if 1,000 persons apply for a job and 500 are initially screened out because they lack the required certificates (pilot or medical), the selection rate in that phase of the pilot selection process is 50%. In the next phase, those 500 applicants take some type of test and just 100 score at or above the “passing” score. The selection rate at that hurdle in the hiring process is .20 (20%). The *Uniform Guidelines* require an employer to request self-identified or self-reported race, sex, and national origin data from applicants to support “applicant flow analyses.” In applicant flow analyses, the selection (pass) rate at each hurdle in the selection process is computed by group. For example, an employer might use a test of aviation knowledge as a hurdle in a pilot selection process. The employer would compute the pass rates for self-identified whites, self-identified blacks (African-Americans), self-identified Hispanic/Latinos, etc. on that test of aviation knowledge as part of the applicant flow analysis. Then the selection (pass) rates would be compared (black to white; Hispanic-Latino to white; etc.) by dividing the minority group selection rate by the majority (e.g., white) selection rate. For example, the pass rate on the aviation knowledge test might be .62 (62%) for black applicants, and .84 (84%) for white applicants. The “adverse impact ratio” is the black pass rate divided by the white pass rate (.62 divided by .84). Under the *Uniform Guidelines*, a selection rate for the protected or minority group “…that is less than fourth-fifths (4/5^{ths}) (or 80% or .80) of the rate for the group with the highest rate will generally be regarded by the Federal enforcement agencies as evidence of adverse impact…” (29 C.F.R. § 1607.4D). In the hypothetical example of the aviation knowledge test, the black selection (pass) rate of .62 is 74% of the white selection rate of .84, indicating that the hypothetical test has adverse impact on black (African-American) applicants. A finding of adverse impact at any stage of the selection process can trigger legal liability for an employer.

The *Uniform Guidelines* and accepted professional practices and principles recommend an evaluation of fairness and the likelihood of adverse impact at each hurdle in a selection process as an element of the overall development and validation effort, *where technically feasible*, prior to implementation of the selection process. Technical feasibility of fairness studies depend on factors such as the number of minority members in the validation sample and other statistical considerations (see 29 C.F.R. § 1607.14C(8)).

It is important to note here that the *Uniform Guidelines* and relevant law *do not* in any way prevent or forbid an employer from using a selection procedure that has adverse impact against one or more minority groups, *provided that the selection procedure is valid*, that is, job-related, and that no alternative with “substantially equal validity” and less adverse impact is available at a reasonable cost.

Best practice #6: Assess selection rates for each demographic group based on cut-scores. Evaluate alternatives such as changing cut-scores, changing weights, predicted job performance, and resulting selection rates by demographic group. Document the rationale for chosen cut-scores and resulting selection rates based on the analysis of alternative scenarios.

Assess selection process utility

Ultimately, the purpose of a personnel selection process for an organization is to select in those individuals that will perform the target job at or above some level of acceptable job performance and to eliminate those individuals who do not have a reasonable likelihood of performing at or above that level.

In personnel selection, this is known as "utility." Utility in this context refers to the usefulness of the selection procedure in achieving organizational goals. Utility can be demonstrated through analyses such as the proportion of selectees succeeding in training and on the job (within a certain period of time after hire) and through financial measures (see Cascio, 1999). Other relevant utility measures include the false positive rate (e.g., the number of persons recommended to hire who go on to fail or perform unacceptably or poorly), costs of development and administration, and selection ratios overall and by demographic group. Constant monitoring of measures like these provides information to senior management about the costs and benefits of the selection process. Unexpected changes in utility can also signal problems with the selection process such as security compromises.

Best practice #7: Assess the utility of the selection process on a regular basis to determine the degree to which organizational goals are being achieved. Monitor the costs of the selection process relative to the financial and operational benefits derived through a reliable and valid selection process.

Challenges in applying these best practices

There are many challenges in applying these best practices in the "real world." Often time is a significant constraint. Funding the work is almost always an issue, for both development and administration of a selection process. Another challenge is accessing and obtaining relevant predictor and job performance data for the validation study. A fourth challenge concerns the trade-offs between validity and equity (Pyburn, Ployhart, & Kravitz, 2008; Sackett, De Corte, & Lievens, 2008). A final challenge is applicant reaction to a selection process. Issues such as perceived validity, procedural fairness, and outcomes can lead to favorable or unfavorable applicant reactions, which in turn, particularly in tight-knit applicant communities such as pilots, can impact the reputation of an employer and its ability to attract qualified candidates (Ryan, & Ployhart, 2000; Truxillo, Bodner, Bertolino, Bauer, & Yonce, 2009). The Society for Human Resource Management (SHRM) summarized these challenges in a guide to selection methods for human resources management practitioners, as shown in Table 1 below.

Table 1: Challenges associated with personnel assessment methods on four key criteria

	Four Key Criteria for Selection Methods				
			Cost to		
Assessment Method	Validity	Adverse Impact (Fairness)	Develop	Administer	Applicant Reactions
Cognitive ability tests	High	High (against minorities)	Low	Low	Somewhat favorable
Job knowledge tests	High	High (against minorities)	Low	Low	More favorable
Personality tests	Low to moderate	Low	Low	Low	Less favorable
Biographical data inventories	Moderate	Low to High for different types	High	Low	Less favorable
Integrity (Honesty) tests	Moderate to high	Low	Low	Low	Less favorable
Structured Interviews	Moderate to high	Low	High	Low	More favorable
Physical Fitness tests	Moderate to high	High (against Females & Older Workers)	High	High	More favorable
Situational Judgment Tests	Moderate	Moderate (against Minorities)	High	Low	More favorable
Work Samples (including Simulation)	High	Low	High	High	More favorable
Assessment Centers	High	Low	High	High	More favorable

Adapted from Pulakos, E. (2005). *Selection assessment methods: A guide to implementing formal assessments to build a high-quality workforce.* Alexandria, VA: SHRM Foundation. Used with permission.

CHAPTER 2: PATHWAYS FOR BECOMING A PILOT

Our next step in the review was to identify and describe the major pathways for becoming a pilot in the United States and other countries. We identified five major pathways: (1) university-based pilot training programs; (2) independent pilot schools/flight academies; (3) the military; (4) Multi-Crew Pilot License (MPL) programs sanctioned by the International Civil Aviation Organization (ICAO, 2006), and (5) "flow-through" agreements in the U.S. between university-based and pilot schools/academies and specific regional and major air carriers.

The FAA regulates these pathways under three "parts" of Title 14 of the Code of Federal Regulations: Part 61, 141, and 142. Part 61 sets the minimum standards for pilot training, including instructor qualifications, flight time requirements, and flight training equipment (aircraft and simulators). Part 61 also sets standards for pilot certification, from student pilot through ATP. Pilot training programs operating under Parts 141 and 142 are regulated more closely by the FAA, including the syllabus used, instructor qualifications, and flight training equipment such as aircraft and flight simulation training devices (e.g., simulators).

Pathway 1: University-based pilot training programs

There are a number of university-based pilot training programs worldwide and in the U.S. For example, the Aviation Accreditation Board International (AABI) lists accredited degree programs offered at 37 post-secondary educational institutions. Example programs are a Bachelor's in Science (B.S.) in Aeronautical Management Technology- Air Transportation Management at Arizona State University, a B.S. in Aviation Management: Flight Option at Florida Institute of Technology, and a Bachelor's of Arts in Social Sciences: Air Transportation with Professional Pilot Specialization at the Ohio State University. In the U.S., these programs generally include the acquisition of an associate's degree or bachelor's degree with the latter being more typical. The aviation degrees often allow the students to focus on general or aviation specific management, or aerospace/aeronautics studies although other degrees can be viable options, depending on the university. The programs generally enable the student to complete the requirements for both an academic degree and initial pilot certification. Certification of the pilot school is usually under Part 141 (e.g., 14 C.F.R. § 141), allowing the students to complete each phase of the training with fewer flying hours than is possible in a Part 61 program (e.g. as few as 190 hours minimum for a Commercial Pilot (CP) certificate under Part 141 versus 250 hours under Part 61). Part 141 pilot school certification requires submission of the curriculum for approval by the FAA. Part 141 also imposes personnel requirements on the school, such as the qualifications of the chief flight instructor. FAA also provides a greater degree of regulatory oversight to Part 141 pilot schools. For example, the curriculum offered by a Part 141 program must meet specific FAA criteria and is periodically reviewed by the FAA. Importantly, however, the same performance standards for a pilot certificate or rating are required under both Parts 61 and 141.

Entry into a Part 141 collegiate aviation program in the U.S. depends largely on the academic admissions criteria used by the college or university. For example, admission to the University of Oklahoma is based on (a) completion of high school classes in English, math, science, and history, (b) high school grade point average (GPA), and (c) standardized test score (SAT® or ACT®).. There are no special aviation-specific requirements to *enter* into the aviation program and pilot training. "Special permission" is required for every flying course, where special permission means essentially the instructor's permission. In addition, the student pilot must pay an additional fee for flight instruction on top of regular tuition. The student pilot must maintain a satisfactory GPA and complete certain required business and mathematics courses to be retained in the program. Graduation requires completion of all degree requirements of the University. This is typical of other public and private university-based collegiate pilot training programs.

Pathway 2: Independent flight schools/flight academies

A person aspiring to become a pilot can also seek pilot training outside of a traditional post-secondary educational (college) setting. Choices include receiving training from an individual certified flight instructor, flight training offered by a fixed-base operator (FBO), pilot training programs unaffiliated with a college or university operating under either Part 61 or Part 141, and pilot training centers operating under Part 142. As noted in the discussion of university-based pilot training programs, there are differences in training structure, minimum flight hour requirements, and degree of FAA oversight between programs operating under Part 61 versus Part 141 and 142.

These schools typically provide ground instruction in preparation for the written exams combined with flight training. The certificates and ratings available from these flight schools can vary and is dependent on the aircraft available and instructors on staff. For example, some smaller operations may not offer training for the multi-engine rating if a multi-engine aircraft is unavailable or if a multi-engine instructor is not on the staff. Pilots that opt for independent pilot school training can, and often do, earn certificates and ratings at different schools based on a number of factors including personal finances as well as instructor and equipment availability.

While the FAA does maintain a list of pilot schools (http://av-info.faa.gov/PilotSchool.asp), there is no central clearinghouse of information on *admissions criteria* for these programs. We searched for ATP training programs on the internet and reviewed admissions information for a number of independent pilot schools/flight academies in the U.S. We also searched the National Center for Educational Statistics (http://nces.ed.gov) for programs coded as providing *Airline/Commercial/Professional Pilot and Flight Crew* (Classification of Instructional Programs code 49.0102) not associated with a university or college. Generally speaking, professional pilot-oriented programs appear to require at least a Student Pilot certificate and valid first-class medical certificate. Other criteria included previous academic work at the college level (if applicable) and, apparently at some schools, "acceptable" scores on the SAT or ACT college admissions tests. Minimum SAT or ACT score was not published on the websites for any program. Financial costs are significant for these independent flight schools/flight academies, running into the tens of thousands of dollars. General estimates range from $30,000 on the low end to upwards of $70,000 or more (Hilkevitch, 2013). These programs generally take less time than university-based programs, requiring anywhere from 90 to 180 days (3 to 6 months).

We also reviewed the entry criteria for several non- U.S. flight academies. Those results are summarized in Appendix B by program. Generally speaking, completion of a secondary (high school) education is required, sometimes with a minimum GPA (or equivalent). Many have a requirement for course work in English, Physics, and Mathematics, some with a minimum grade requirement. English proficiency (both written and spoken) is another common requirement for entry into the flight academies. Most programs appear to have a minimum age; some also have a maximum age limit for entry.

Pathway 3: Former military pilot

Former military pilots were traditionally the main source for airline pilots but this situation has been changing steadily since the 1980s (Collup, 2007). Military pilots are the product of a stringent selection and training environment resulting in severe range restriction when evaluating selection procedures. Military pilots typically participate in primary and advanced flight training programs that provide amounts of flying experience roughly equivalent to a civilian commercial pilot certificate. The generally more demanding performance requirements in military flight training and operations are acknowledged with the current restricted ATP that is available to former military pilots with 750 hours of flight time.

U.S. Air Force, U.S. Navy, and U.S. Marine Corps pilots are required to have a college (undergraduate) degree and are commissioned officers. U.S. Army pilots are not required to have college degrees. Selection into military flight training has been extensively researched and documented (see Damos, 2011). Historically, U.S. air carriers preferred former military pilots over other sources. However, there are far fewer former military pilots in recent years than in past decades, and the carriers have turned to other sources. According to FAA testimony, upwards of 80% of airline pilots had prior military pilot experience through the mid-1990s (*The Age 60 Law*, 2001). But by the early 2000s, just 40% of newly hired airline pilots were from the military. Of those pilots hired from 2005 through 2009 by six regional carriers, only 3% were former military pilots (Smith, Bjerke, NewMeyer, Niemczyk, & Hamilton, 2010). In a follow-on study of pilot hiring in 2013 through 2015, about 12% of pilots hired by 22 regional carriers were former military pilots (Bjerke, Smith, Smith, Christensen, Carney, Craig, & Niemczyk, 2016). Even with that recent upswing, fewer former military pilots have been hired by U.S. air carriers in recent years than in the 1990s.

Pathway 4: Multi-Crew Pilot License (MPL)

Pilot training in the previous pathways is largely based on prescriptive task-oriented and hours-based syllabi (International Air Transport Association (IATA), 2011). However, beginning in the late 1990s, calls for review and adjustment of pilot training and licensing methods began to emerge. The International Civil Aviation Organization (ICAO) convened a Flight Crew Licensing and Training Panel (FCLTP) in 2001. Increasing automation on the flight deck was thought to have resulted in a shift of required skills from manual flying skills to "system operator skills" (Scheck, 2006, p. 21). Moreover, advances such as "competency-based training," instructional systems design (ISD), and ever-more advanced and capable Flight Simulation Training Devices (FSTDs) might be exploited to modernize pilot training. The end result was development of the Multi-Crew Pilot License (MPL; ICAO, 2006, p. 3-1) concept.

The design of MPL-compliant training programs arguably provides student pilots with an experience level comparable to traditional training programs but with more time spent in the simulator environment. This is combined with fewer hours in an aircraft and more time training in simulated situations that emulate operations in a multi-pilot aircraft with scenario complexities that reflect operational line flying. Total hours at the completion of an MPL program range from 240 to 325 hours with 80 to 100 of those hours being in the aircraft and the balance in a simulator (Boudreau, 2014). As with traditional training programs, MPL programs begin with a focus on basic flying skills, instrument training, and simulator-based line-oriented flight training (LOFT) with scenarios representative of situations encountered in "line" or actual flight operations. However, students in MPL programs are typically exposed to more complex aircraft earlier in the training based on the concept that more time training in the types of aircraft they will be operating during their careers will better prepare them and improve transfer of training.

MPL programs are in use around the world. According to the IATA "Global MPL Course Tracker" (http://www.iata.org/whatwedo/ops-infra/itqi/Documents/MPL%20documents/mpl-courses-global-tracker.xls). As of October 2017, there were 39 MPL programs in the IATA database, with 2,643 graduates and 4,669 MPL students. Course lengths range from 14 to 24 months, with 246 to 404 hours of flight instruction and 255 to 292 hours in an FSTD. Interestingly, flight hours in an actual aircraft ranged from just 79 to 115 flight hours, with 10 to 30 solo hours and 6 to 20 landings. The single largest MPL program, according to the IATA on-line data, is Lufthansa (Lufthansa, German Wings International, Lufthansa Flight Training, Lufthansa Cityline) with 1,144 graduates and 1,488 students as of October 2017. MPL graduates are initially restricted to serving as co-pilot until meeting the regulatory and flight hour requirements for an unrestricted ATP certificate issued by their national aviation authority.

MPL programs focus on the nine overall competencies: 1) application of procedures; 2) communication; 3) flight management, guidance, and automation; 4) manual aircraft control; 5) leadership and teamwork; 6) problem solving and decision making; 7) situation awareness; 8) workload management; and 9) (technical) knowledge. These competencies are assessed during all phases of flight from ground operations through landing. While the MPL approach appears to be gaining traction around the world, concerns are emerging about the conceptual model and practical application. For example, Wikander and Dahlstrom (2014, 2016) identified a number of concerns with MPL programs such as the lack of research to support the assumptions on which the program is based, differences in competencies assessed across training programs, and (in)consistent measurement of the competencies. The number of students enrolled in each of 31 MPL programs is reported in the IATA global MPL tracker, as is the number of graduates for each program. But the number of students still in training and the number of failures are not reported, making it difficult to gauge the overall success rates of the programs.

Note, however, that MPL is *not* currently authorized for training pilots for U.S.-certificated carriers and is *not* a recognized pilot certificate for U.S. pilots. Foreign pilots with the MPL, serving as a co-pilot, can fly into the U.S., provided the aircraft is not a U.S.-registered aircraft. However, foreign pilots with only an MPL cannot operate a U.S.-registered aircraft for a U.S. carrier. Therefore, MPL is *not* a pathway for U.S. pilots into the cockpit of U.S. air carriers. However, foreign operators do send MPL students to U.S. pilot training schools for flight training.

Pathway 5: "Flow-through" agreements

"Flow-through" agreements are an emerging pathway to becoming a pilot for a Part 121-certificated carrier in the United States. These are also known as "pilot pipeline" agreements. We identified three variations in these flow-through agreements: a) university-regional-major agreements; 2) flight school-regional-major agreements; and 3) regional-major agreements. In the first variation, a university-based pilot training program partners with a regional carrier. For example, as of March 2018, Envoy Air® has partnerships with 22 collegiate-based pilot training programs and six flight academies (https://www.envoyair.com/cadet-program/). To qualify for the “cadet” program, the student must hold a Private Pilot certificate and Instrument rating and have a 2.50 or higher cumulative grade point average overall and 3.00 or higher in aviation coursework. The program begins with an interview by Envoy Air® in the student’s sophomore year. The student accumulates flight hours while completing the undergraduate degree requirements and also earns the Commercial Pilot certificate with Multi-engine aircraft and instrument ratings. The graduate then works as a Certified Flight Instructor at the pipeline university to build hours toward the restricted ATP certificate. Envoy Air® indicates that a technical and simulator evaluation is also required as part of the overall “interview” process. But the most important aspect of the program is that once the graduate attains enough hours to qualify for the ATP certificate, he or she can “flow through to American Airlines with no additional interview required.”

An agreement between EagleJet International and PSA Airlines illustrates the second variation in flow-through agreements between a non-academic pilot training program and regional carrier (http://www.eaglejet.net/index.asp). The program offers a “guaranteed interview with PSA Airlines prior to starting” the EagleJet program. Success in that “guaranteed interview” appears to be a criterion for selection into the EagleJet-PSA program, along with the ability to pay $29,000 for the training. Similarly, the program appears to offer a “guaranteed” conditional offer of employment with PSA Airlines, conditioned on successful completion of the EagleJet program and a “successful interview” with PSA. It is not clear if this is a second post-entry interview with PSA. A Commercial Pilot certificate, Class I medical certificate, multi-engine aircraft and instrument ratings, a Federal Communications Commission (FCC) restricted radiotelephone operator’s license1 (required to operate the aircraft radio), a valid driver’s license, and U.S. citizenship (or if foreign, authorization to work in the U.S. and “perfect English skills”) are required to enter the EagleJet program. Financing for the program fee ($29,000) is available. The student “flows through to the right seat of a PSA regional jet” once 1,500 flight hours are accumulated and the ATP certificate earned. It is not clear if there are any other criteria for selection into the “right seat of a PSA regional jet” other than an interview.

CHAPTER 3: CIVILIAN PILOT SELECTION

Having identified the four major pathways for becoming a pilot for a U.S.-certificated air carrier, we next turned our attention to identifying the various hurdles embedded in the pathways. By "hurdles," we mean the criteria, requirements, assessments, tests, examinations, and simulations that a person must meet, present, or pass to become a pilot, from initial application to a training school through selection into a major airline's pilot training program. This chapter is organized as follows.

First, we briefly summarize previous reviews of pilot selection, focusing on civilian pilot selection. Second, we consider the characteristics of pilot aptitude test batteries we identified in the review of selection procedures for operators around the world. In particular, we searched for technical information on the test batteries such as constructs measured and psychometric characteristics such as reliability. We also searched for information describing how the test batteries were validated in relation to pilot performance. We evaluated each test battery relative to the "best practices" framework we developed on the basis of current U.S. legal guidelines, case law, and professional standards, principles, and practices. In the third section of the chapter, we consider assessments of non-technical interpersonal competencies or abilities in the realm of CRM. We consider interviews in the final section of this chapter. We focused on civilian pilots and the "select in" phase of hiring exclusively throughout our search and review. We do *not* consider post-offer/post-employment "select out" processes such as training, medical examinations, or mental health assessments in which the intent is to remove a pilot from employment.

Previous reviews of (civilian) pilot selection practices

Researchers have conducted several extensive reviews of pilot selection procedures with some within the past 25 years being arguably the most relevant (Air Line Pilot's Association, International, 2009; Damos, 1996, 2003, 2011; Hunter & Burke, 1994, 1995; Young, Schroeder, Agen, & Broach, 1998). The literature on pilot selection has also informed guidance on best practices for developing pilot selection procedures (IATA, 2012). Common features of pilot selection systems include age, educational accomplishment, English proficiency (outside of English-speaking countries), and often, successful completion of course work in mathematics and English. Interviews are very common as well. Caution should be taken as most of the research and the associated reviews of pilot selection involved military pilots with minimal or no evaluation of pilot selection and training in the civilian sector. Empirical questions remain concerning differences that may exist between those who opt for military or civilian training, the nature of the training, the flight environments, and criterion measures appropriate to each environment.

The 2012 report on an IATA survey of airlines came to this dismal conclusion: "Despite the clear benefits of a proper pilot selection process, the results showed that only a minority of airlines have a specific selection system in place that is structured and scientifically based" (p. 1). IATA also noted that "Current [pilot] selection systems appeared to lack a conceptual basis – there is a need for conceptual support in setting up an efficient selection system" (pg. 10). Finally, IATA observed that "Aptitude testing plays the key role in the recruitment process because of its importance for the quality of hired staff" (pg. 18).

Review of pilot aptitude test batteries

Aptitude testing in many forms has been a mainstay of military and civilian *ab initio* pilot selection since the early 1900s. There is over 100 years of research supporting the use of psychometric aptitude testing for pilots. While extensive documentation on U.S. military pilot selection aptitude testing has been published by the U.S. Air Force and U.S. Navy, in particular, civilian pilot selection aptitude testing is less accessible. We conducted a literature search for civilian pilot selection research in multiple databases and via Google Scholar®. We also conducted general Internet searches using key phrases such as "pilot aptitude test" and "pilot selection." In particular, we searched for descriptions of pilot selection processes and technical information on those procedures such as validation studies. We also contacted researchers active in the domain via e-mail, requesting information. We focused in particular on aptitude tests or test batteries.

We identified 15 pilot aptitude tests or test batteries available in the global marketplace. Many of these test batteries are in use overseas, either by pilot training organizations or by air carriers. The test batteries include the *Computerized Pilot Aptitude Screening System* (*COMPASS*®; European Pilot Selection & Training, 2015), *PILAPT*® (People Technologies, Inc., 2015), Gessellschaft für Angewandte Psychogische Forschung (*GAPF*; 2015), WOMBAT and WOMBAT-FC (Roscoe, Corl, & LaRoche, 1997, 1999), and the Deutsche Forschungsanstalt fuer Luft- und Raumfahrt e.V. (DLR, 2015) test battery. Other batteries include *ADAPT* and Future Adaptive Screening Tool (*FAST*), both products of the Resource Group (2015a, b) We used social network analysis tools to develop a graphical representation of test batteries and users. As shown in Figure 1, the *COMPASS* test battery has the largest number of reported users, followed by *PILAPT*. Fewer users were reported for the *GAPF*, *ADAPT*, and *DLR* batteries. The *DLR* test battery is used by one U.S. mainline air carrier, which also uses a customized personality assessment. Another U.S. mainline air carrier uses custom-developed tests and the aeromedical edition of CogScreen® (a neuropsychological test battery; Kay, 1995) in pilot selection.

We then evaluated the available technical documentation for each of the major test batteries against the best practices framework described in the first chapter of this technical report. The results of the evaluation are summarized in Table 2 by test battery. An analysis of a sample of airline selection procedures relative to the best practices is summarized in Table 3. This analysis is restricted to airlines that disclosed details on the tests used.

Best practice #1: Conduct a job analysis. Just two pilot selection test batteries referenced a recent pilot job/task analysis as the basis of the characteristics assessed by the battery. These two batteries were the DLR and a U.S. air carrier test battery. Six of the test batteries referenced the mid-1990s meta-analysis of Hunter and Burke (Hunter & Burke, 1994, 1995) as the basis for the characteristics assessed by the battery. There are two problems with citing the meta-analysis with reference to our best practices framework. First, the Hunter and Burke meta-analysis was based primarily on military pilot selection, not civilian pilots. There are substantial differences in the

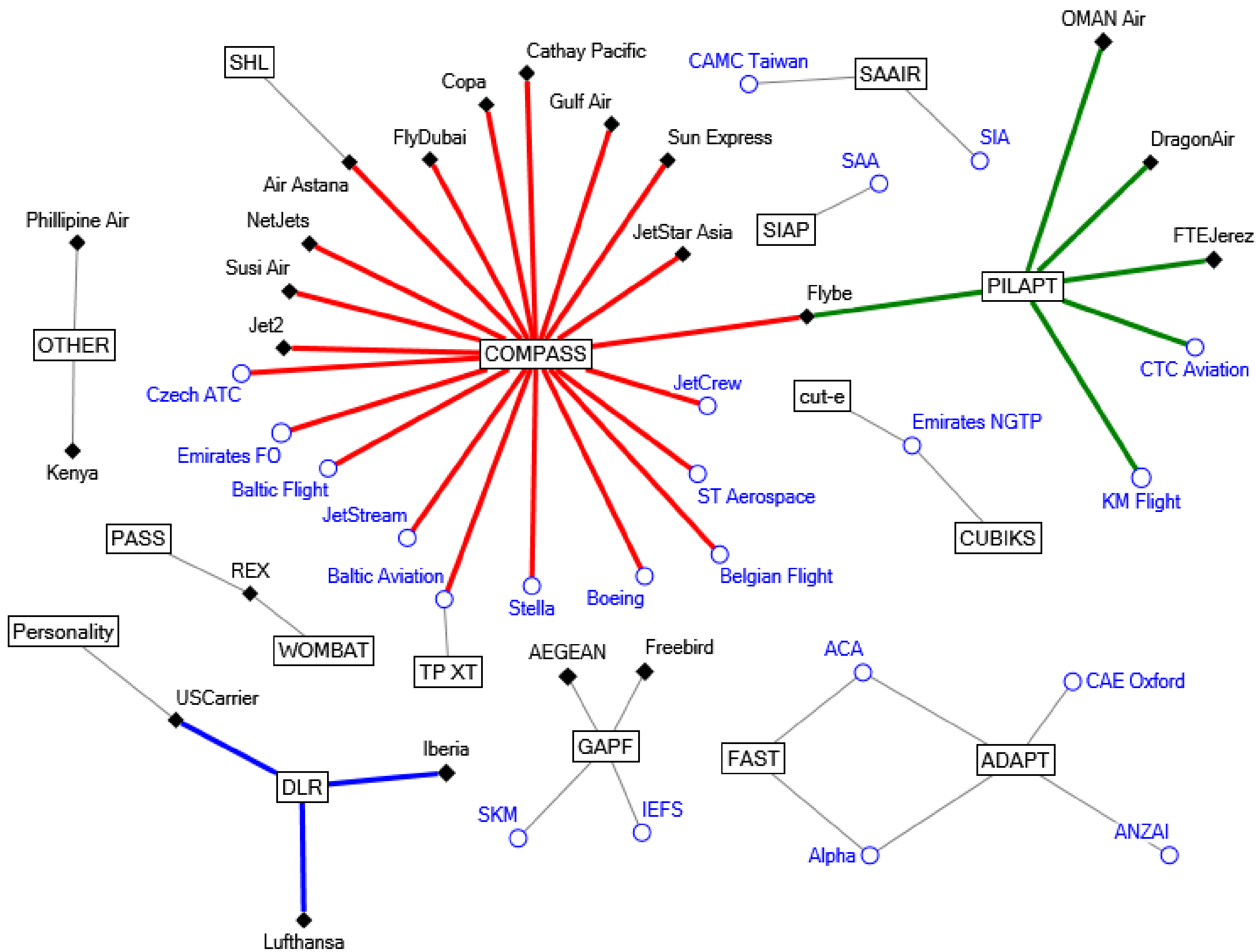

Figure 1: Test-User relationships (Test battery names are boxed, in all caps; Air carrier users are indicated by solid black diamonds; and Flight school/Academy users are indicated by blue open circles). Refer to Appendix C for test battery descriptions.

Table 2: Analysis of pilot aptitude test batteries relative to personnel selection best practices

Best Practice	AC[1]	ADAPT[2]	COMPASS	GPSS[3]	PILAPT[4]	SAAIR[5]	GAPF[6]	SIAP[7]	DLR	WOMBAT
JOB ANALYSIS										
Recent JTA	No	No	No	No	No	No	No	No	**Yes**	No
Meta-Analysis/JAR	N/A			**Yes**	**Yes**	**Yes**	N/A	**Yes**	**Yes**	**Yes**
CRITERIA										
PREDICTORS										
Domain/Construct										
Mathematics	**Yes**	**Yes**	**Yes**			**Yes**	**Yes**	Optional	**Yes**	
Physics		**Yes**	Optional				**Yes**			
Language	**Yes**	**Yes**	Optional	**Yes**			**Yes**	Optional	**Yes**	
Technical	**Yes**	**Yes**	**Yes**				**Yes**	**Yes**	**Yes**	
Instrumentation	**Yes**				**Yes**		**Yes**	**Yes**		
Psychomotor Tasks		**Yes**	**Yes**		**Yes**	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**
Tracking		**Yes**	**Yes**		**Yes**	**Yes**	**Yes**		**Yes**	**Yes**
Attention/Vigilance	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**
Memory	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**
Numerical Reasoning	**Yes**	**Yes**		**Yes**	**Yes**	**Yes**		**Yes**	**Yes**	
Cognitive/Reasoning	**Yes**	**Yes**	Optional	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**
Multi-Task	**Yes**		**Yes**	**Yes**	**Yes**		**Yes**	**Yes**	**Yes**	**Yes**
Spatial/Orientation	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**
Personality	**Yes**	**Yes**	**Yes**	**Yes**			**Yes**	**Yes**	**Yes**	Optional
Predictor Psychometrics										

Reliability	N/A	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**	N/A	**Yes**	**Yes**	N/A
Construct Validity	N/A	**Yes**	N/A	N/A	**Yes**	**Yes**	N/A			**Yes**
Testing Method[8]										
Group Task/Teamwork	**Yes**						**Yes**	**Yes**	**Yes**	
Assessment Center	**Yes**								**Yes**	
Interview	**Yes**	**Yes**					**Yes**	**Yes**	**Yes**	Optional
Simulation		**Yes**					**Yes**		**Yes**	
VALIDATION										
Criterion Validation	N/A	N/A	N/A	N/A	**Yes**	**Yes**	**Yes**	Yes	**Yes**	**Yes**
Content Validation	N/A	N/A	N/A	N/A	**Yes**	**Yes**	**Yes**	Yes	**Yes**	N/A
FAIRNESS										
Adverse Impact	N/A	N/A	N/A	N/A	N/A	**Yes**	N/A	N/A	N/A	N/A

Notes: [1]AC: Albert Consulting (precise details on test not provided on website). [2]ADAPT: 3 versions of test battery published by Symbiotics, Ltd. [3]GPSS: 14 tests in battery; several appear to assess similar or overall dimensions. [4]PILAPT: 8 tests in battery; some involve overlapping dimensions. [5]SAAIR: Vienna Test System. [6]GAPF: Society for Applied Psychological Research. [7]SIAP: Scandinavian Institute for Aviation Psychology. [8]All test batteries feature tests taken by an individual working alone.

Table 3: Analysis of test batteries used and disclosed by operators or training programs relative to best practices in personnel selection (based on publicly available information)

Best Practice	KLM	Delta	ENAC[1]	Quantas[2]	SAS	Thai Airways[3]	Cathay Pacific[4]
JOB ANALYSIS							
Recent JTA	No	**Yes**		No	N/A	No	**Yes**
Meta-Analysis/JAR	**Yes**	**Yes**		**Yes**	N/A	**Yes**	
CRITERIA							
PREDICTORS							
Domain/Construct							
Mathematics	**Yes**		**Yes**			**Yes**	Yes
Physics			**Yes**			**Yes**	Optional
Language			**Yes**			**Yes**	
Technical	**Yes**	**Yes**				**Yes**	**Yes**
Instrumentation				**Yes**		**Yes**	**Yes**
Psychomotor Tasks				**Yes**		**Yes**	**Yes**
Tracking			**Yes**				**Yes**
Attention/Vigilance		**Yes**	**Yes**			**Yes**	**Yes**
Memory		**Yes**	**Yes**	**Yes**		**Yes**	**Yes**
Numerical Reasoning		**Yes**	**Yes**	**Yes**		**Yes**	
Cognitive/Reasoning	**Yes**	**Yes**	**Yes**	**Yes**		**Yes**	**Yes**
Multi-Task	**Yes**	**Yes**	**Yes**	**Yes**		**Yes**	**Yes**
Spatial/Orientation	**Yes**	**Yes**	**Yes**	**Yes**		**Yes**	**Yes**
Personality	**Yes**	**Yes**		**Yes**	**Yes**	**Yes**	**Yes**
Predictor Psychometrics							
Reliability	N/A	**Yes**	**Yes**	N/A	N/A	**Yes**	**Yes**
Construct Validity	N/A			N/A	N/A		
Testing Method							
Group Task/Teamwork			**Yes**		Yes	Yes	Yes[5]
Assessment Center							
Interview	**Yes**	**Yes**	**Yes**	Yes	**Yes**	**Yes**	**Yes**
Simulation	**Yes**				**Yes**		Yes[5]
VALIDATION							

Content Validation	N/A	**Yes**		N/A	N/A	**Yes**	
Criterion Validation	**Yes**	**Yes**	In process	N/A	N/A	**N/A**	N/A
FAIRNESS							
Adverse Impact	N/A		N/A	N/A	N/A	N/A	N/A

Notes: [1]ENAC: Ecole Nationale d'Aviation Civile, France (see Doat, 1995). [2]Quantas Cadet Pilot program (Stead, 1991a, 1991b). [3]Thai Airways uses SIAP. [4]Cathay Pacific uses COMPASS (Bartram & Baxter, 1996). [5]Group Task/Teamwork and Simulation used by Cathay Pacific for advanced entrants only (not for *ab initio* candidates)

actual flying and in the organizational context for military and civilian pilots. The likelihood that one can generalize from military to civilian requirements is not well established. Second, a meta-analysis of test validities is likely to be indefensible in the U.S. Selection procedures with adverse impact relying on validity generalization have not fared well in litigation in the U.S., particularly when there has not been a supporting job analysis (Biddle, 2010; Tonowski, 2015).

Best practice #2: Define job performance measures. Little attention has been given to the measures of job performance against which the pilot selection test batteries are validated (Broach, 2014). The German DLR has conducted informative research on the assessment of non-technical, largely interpersonal, CRM-related aspects of performance in the cockpit (Flin, Martin, Goeters, Hörmann, Amalberti, Valot, & Nijhuis, 2003; Goeters, 2004). The default criterion appears to be training outcomes at various stages, including *ab initio*, simulator, and what is known as "initial operating experience" (IOE). In IOE, the newly hired pilot is paired with an instructor pilot in operational, revenue flights. The instructor pilot evaluates the new pilot's performance in 25 to 40 hours of actual flight operations.

Best practice #3: Identify predictor measures. As noted in the overview of best practices, a key step in selecting predictor tests is to establish, through a job analysis, what attributes are to be assessed. Only the DLR test battery is based on a current pilot job analysis. The GPSS, PILAPT, SAAIR, SIAP, and WOMBAT batteries are based on previously published meta-analyses (particularly the Hunter & Burke, 1994 meta-analysis). No information about was available for the ADAPT or COMPASS test batteries.

That being said, it appears that pilot test batteries all include assessments of general intelligence ("*g*"). Tests of logical reasoning and verbal, numerical, and spatial abilities are common in pilot test batteries, as shown in Table 2. These cognitive ability tests are typical measure of "*g*" or "general mental ability" (Pulakos, 2005). Attention/vigilance tests are also common, as are tests of "memory," "situational awareness," and "multi-tasking." Some pilot test batteries, as shown in Table 2, also include content-oriented tests in domains such as physics and mathematics. Test batteries have broadened over the last 20 years to include non-cognitive factors such as personality and social skills or competencies. These non-cognitive measures are discussed in more detail in the fourth section of this chapter.

Best practice #4: Evaluate relationship of KSAO measures (predictors) to job performance criteria. Validity, in personnel selection, relates to the degree to which a justifiable inference about future job performance can be drawn on the basis of scores on the tests comprising the test battery. That inference is commonly based on the statistical correlation between test scores and measures of training performance, training outcomes, and/or job performance. Validity data were available for four of nine test batteries. Overall, the validities ranged from .27 to .77 across tests and a variety of training and job performance criteria for samples ranging in size from 32 to 1,760. Seven of the reported studies were based on military pilots and seven on civilian pilots.

Other validity studies were located for these and other test batteries (See Appendix B). However, reporting was highly variable across reports. For example, Sommer, Olbrich, and Arendasy (2004) and Arendasy, Sommer, and Hergovich (2007) reported the accuracy of the PILAPT test battery in predicting positive and negative training outcomes (e.g., true positives and true negatives) but did not report an overall validity coefficient. Hoffman, Spetz, and Hoffman (2011) reported the average number of training problems and lost productivity hours per pilot by quintile of "predicted performance category" but no overall simple or multiple correlation between their pilot selection test battery (CogScreen® AE, NEO-PI(R), and a proprietary Job Knowledge Test) and job performance immediately after hire.

Best practice #5: Set cut-scores and/or decision rules. There is no public information on cut-scores on pilot aptitude test batteries, including the methods used to set the cut-scores. There is also no information on decision rules used by airlines to select and reject applicants on the

Table 4: Validity studies for pilot selection test batteries

Test Battery	Citation	*N*	Group	Statistic	Validity
PILAPT	Burke, Kitching & Valsler (2001)	165	Military	R^2	.56
	Burke, Hobson, & Linksy (1997)	1,760	Military	*r*	.35, .45
	Mesorosova (2013)	221	Civilian	*r*	-.19 - -.25*
	Kokorian (2010), Kokorian, Valsler, Tobar, & Riberio (2004)	67	Military	*r*	.36 - .37
		90	Military	*r*	.36 - .50
		117	Military	*r*	.27
		165	Military	*r*	.26 - .55
		76	Civilian	*r*	.28 - .40
		224	Military	*r*	.08 - .45
SAAIR	Sommer, Häusler, Koning, & Arnedasy (2006)	90	Civilian	R^2	.61
WOMBAT	Frey, Thomas, Walton & Wheeler (2001)	32	Civilian	*r*	.36
DLR	Damitz, Manzey, Kleinmann, & Severin (2003)	1,036	Civilian	*r*	.28
	Hörmann & Luo (1999)	125	Civilian	R^2	.47 - .65
	Stahlberg & Hörmann (1993)	51 - 98	Civilian	R^2	.54 - .77
	Zierke (2014)	402	Civilian	*R*	.14 - .19
SIAP	Syversen (2012)	178	Civilian	*r*	.22 - .32
	Wingestad (2005)	100	Civilian	*r*	.33 - .41

Notes: Mesarosova criterion was extra flight hours required (resulting in negative sign) basis of test scores. The actual cut-scores and decision rules used are confidential with good reason. If such cut-scores and decision rules were publicly known, applicants could (would!) attempt to “game” the tests and hiring processes.

Best practice #6: Evaluate the fairness of the selection procedures. Just one of the nine test batteries has disclosed information on fairness, that is, differences in score distributions by race, sex, national origin, and other demographic factors. **Best practice #7: Assess selection procedure utility.** There is one published evaluation of the utility of a pilot selection process (Goeters & Mashcke, 2004). This is not to say that such evaluations have not been conducted; they are just not available for review, most likely due to competitive pressures.

Innovations in pilot aptitude assessments

The primary innovation in pilot aptitude assessments has been in the incorporation of personality and social skills or competencies into the selection process. Personality in this context refers to normal personality and the prediction of job performance, not assessment for clinical or diagnostic purposes. The assessment of social skills or competencies grew out of the Cockpit Resource Management (CRM) movement and the appreciation of the role and impact of interpersonal exchanges and dynamics in the cockpit on the safe operation of increasingly complex aircraft.

Personality assessment

Efforts to identify personality characteristics that are important factors in predicting pilot performance during training or success in training have generally been of limited success. In their meta-analytic study, Hunter and Burke (1994) found that personality inventories had limited predictive value when compared with cognitive tests, psychomotor tests, and biographical data inventories. Martinussen (1996, 2016) and Damos and Koonce (1997) expressed a similar viewpoint. However, they still point to the importance of developing improved personality measures and use of more effective criterion data to evaluate those measures. In his review of the use of personality measures to evaluate pilot applicants, Maschke (2004) initially identified some of the difficulties associated with the use of those measures and then indicated the variety of measures used to assess applicant personality measures. The difficulties involved in the use of personality measures include the generally lowered reliability when compared with cognitive measures, the criticality of identifying appropriate criterion measures, the non-linearity between predictors and the criteria, cultural differences, and efforts of applicants to present a positive self-image when responding (faking good).

Social skills or competencies

In previous pilot selection research, the search for predictors focused on the importance of crew resource management (CRM) as a factor in aviation accidents/incidents. This line of research can be traced back more than three decades ago to the work of Cooper, White, and Lauber (1979) who, in their review of 10 years of commercial aviation accidents found that the events were more closely associated with breakdowns in communication and coordination than lack of knowledge or technical skill. CRM was popularized by Helmreich and his associates and students and has become embedded in the aviation (and other) industries.

CRM was originally conceptualized as an instructional strategy to train crews in the effective use of available resources in the cockpit (Helmreich, Merritt, & Wilhelm, 1999). CRM has evolved, however, over three decades; Helmreich, Merritt, and Wilhelm (1999) describe the transition from the first generation focus on managerial effectiveness to that of the fifth generation that focused on error management. Maurino and Murray (2010) provide additional insights into this transition. Helmreich, Wilhelm, Klinect, and Merritt (2001) describe the potential influence of the professional cultures of the pilots, the cultures of organizations, and the national cultures and their potential influence on threat and error management and aviation safety. Westrum and Adamski (2010) review some of the research regarding how organizational and cultural factors influence communication and safety in airlines. They reveal that there are differences among nations as well as within nations when you observe command structure, information sharing, and organizational climate within air carriers.

Early research by Helmreich and Foushee (1993) defined some of the team process variables that were important elements to consider for CRM training. Prince and Salas (1999) point out the similarities across research settings regarding six of the seven skill areas they defined (situation awareness, communication, mission analysis or planning, assertiveness, decision making, and leadership). The additional skill set involved adaptability. They go on to describe efforts to measure those dimensions and then develop team training strategies. Significant effort has been invested in developing measures of CRM behaviors in the cockpit (Flin, Goeters, Hörmann, & Martin, 1998; Flin & Martin, 2001; Gontar & Hörmann, 2014; Hedge, Bruskiewicz, Borman, Hanson, Logan, & Siem, 2000; Hörmann, 2004; Hörmann, Burger, & Neb, 2003). CRM behaviors assessed in these studies include communication, problem-solving, decision making, interpersonal skills, situation awareness, and leadership (Franz, Prince, Cannon-Bowers, & Salas, 1990; Helmreich & Foushee, 1993; Salas, Fowlkes, Stout, Milanovich, & Prince, 1999). These measures might serve as both selection and performance criteria and provide a basis for feedback to pilots in training.

Despite the ubiquity of CRM in the world of aviation, its impact on operations and safety is uncertain. Salas, Wilson, Burke, and Wightman (2006) point out that evaluations of the effectiveness of CRM training are often limited and lacking in thoroughness and that there is little evidence that CRM has impacted overall organizational safety (Dahlstrom, Laursen, & Bergstrom, 2008). However, this lack of evidence might be attributed to difficulties such as a lack of resources (O'Connor, Flin, Fletcher, & Hemsley, 2002).

Regardless of the merits of CRM *per se*, the concept led to an increased emphasis on identifying personality and other social competencies that may be used as part of the selection process to improve communication and coordination in the cockpit (Hörmann & Goerke, 2014; Martinussen, 2014). For example, older work by Chidester, Helmreich, Gregorich, and Geis (1991) used three personality scales (Extended Personality Attributes Questionnaire – EPAQ, Work and Family Orientation Questionnaire – WOFO, and Achievement and Impatience Scales – Pred) to identify personality clusters that could influence crew communication and coordination. Chidester, Kanki, Foushee, Dickinson, and Bowles (1990) demonstrated, using full-mission simulations that crews who had a Positive Instrumental/Expressive captain were consistently more effective and made fewer errors than crews led by low motivation captains. However, the track record for personality assessment in pilot selection has not been encouraging. For example, the meta-analytic estimate for the predictive validity of personality tests is .14 (Martinussen, 2014).

More recently, attention has shifted from personality *per se* to social competence (Hörmann & Goerke, 2014). IATA (2012) and others (Hörmann & Maschke, 1996; Goeters, Maschke, & Eißfeldt, 2004) have pointed to the importance of social competency dimensions in assessing flight crew performance. Hörmann and Goerke (2014) cite a presentation that reported results from a safety survey of more than 2,000 Lufthansa pilots demonstrating that "… social factors in the cockpit play a greater role in the frequency, risk, and mastery of safety-related incidents than either human error or operational and technical issues. (p. 7)."

In another investigation of social competence, Hörmann and Goerke (2014) administered two social competence questionnaires to 292 participants in a five-stage pilot selection process. The two questionnaires were a 40 item German version of the *Interpersonal Competence Questionnaire* (ICC) and the 90 item German translation of the *Social Skills Inventory* (SSI). Additional elements of the selection process included the 183 item *Temperament Structure Scales* (TSS; Maschke, 1987), basic cognitive ability tests, psychomotor tests, an assessment center (described below), a simulator run, and an interview. There was evidence of moderate to high correlations between some dimensions of the ICC and SSI and the TSS scales. Neither the ICQ nor SSI contributed significantly to the selection outcomes when included with the other variables derived from the five-stage selection process. With respect to predicting training success, the better predictors involved the cognitive and psychomotor performance tests and the work sample simulator test. Longer term investigations are needed to determine if any of the social competence measures actually predict CRM and other relevant behaviors in the cockpit.

Assessment Centers

Another emerging approach is to the use of an Assessment Center (AC). The AC approach focuses on candidate behavior in exercises with evaluation by trained observers ("assessors") on specific dimensions (or competencies) (Arthur & Day, 2011). An example of a well-documented pilot candidate assessment center is the Lufthansa pilot selection process as administered by the DLR, the German national aviation research laboratory in Hamburg (Hörmann, Manzey, Maschke, & Pecena, 1997; Damitz, Manzey, Kleinmann, & Severin, 2003; Hörmann & Goerke, 2014). The DLR AC is the third of five stages in pilot selection. The stages are (a) basic pilot aptitude tests (36% pass rate); (b) psychomotor tests (87% pass rate); (c) the AC with four exercises (65% pass rate); (d) a fixed-base flight simulator (77% pass rate); and (e) an interview (73% pass rate). The overall selection (pass) rate across the five stages was about 11% of 305 applicants for *ab initio* training with Lufthansa (Hörmann & Goerke).

The four AC exercises focused on the goal-oriented social behavior of the pilot applicants. The exercises include two problem-solving tasks in small groups, a role-playing exercise to resolve a conflict, and a test of cooperation in which two candidates worked together via two inter-connected computers. Candidates were assessed on seven behavioral dimensions in the course of these exercises by four trained observers. With respect to pilot AC construct validity, Damitz, Manzey, Kleinmann, and Severin (2003) derived two factors – performance competence and interpersonal competence – from the seven rating dimensions across the four exercises. The performance-related competence ratings had a higher correlation with the *g*-component (general intelligence) derived from the basic and psychomotor ability measures (r = .23) than the interpersonal competence rating (r = .17)). When corrected for restriction in range, the overall correlations between factor scores and the criterion measure (peer ratings) were .27 for interpersonal competence and .36 for performance competence. Damitz et al. concluded that "…the assessment center approach represents a promising tool for assessing personality-related qualities of (student) pilots which determine their interactions with others and their efficiency in teamwork situations (p. 209)." As the authors indicate there is a need to obtain additional criterion-related data.

Interviews

Interviews appear to be a ubiquitous feature of civilian pilot selection programs. For examples, interviews are an element of the selection process for CAE pilot training programs conducted on behalf of carriers such as Tigerair® and EasyJet®, the KLM Flight Academy, FTE Jerez, and U.S. air carriers. A Google™ search on "airline pilot interviews" returned about 728,000 results with titles such as "*Pilot interview information gouge*" (where "gouge" is a colloquial expression for "latest information" on a topic) and links to books such as *Airline pilot interviews: How you can succeed in getting hired*. A similar search on "airline pilot interview questions" returned about 185,000 results with titles such as "*836 FREE pilot interview study guide questions and answers*" and "*Airline pilot interview: 10 questions you will be asked*."

There are five primary issues with interviews as personnel selection procedures: 1) what is assessed in the interview, 2) how the interview is conducted, 3) how the interview is scored, 4) when (in the hiring process) the interview is conducted, and 5) how the results are used in the employment process. The first issue is what is assessed in the interview, that is, what psychological constructs are measured through the interview. An interview might be intended to assess "motivation" and "cultural fit" or other "soft" competencies. An interview might also be used to determine a candidate's familiarity with the systems and controls of a specific class or type of aircraft, that is, specific technical knowledge. The range of constructs that might be assessed in an interview is quite large. However, the primary focus of interviews appears to be on non-technical constructs such as person-organization fit, motivation, personality, and social competencies. At least one European air carrier conducts a clinical interview, to assess the psychological-emotional status of candidates.

The second issue concerns how the interview is conducted. The "how" of an interview concerns factor such as the types of questions asked, the degree to which an interview is structured or unstructured, whether it is conducted by a single individual or a panel, and the medium in which the interview is conducted (e.g., face-to-face, telephone, or videoconference). Three common interview question types are factual, past behavior, and situational behavior (future behavior) (Campion, Campion, & Hudson, 1994). With regard to structure, strong empirical evidence indicates that structured interviews are more valid predictors of future job performance than what IATA (2012) calls "free style" unstructured interviews (Dipboye, 1994; Huffcut, 2011; Huffcut & Culbertson, 2010). A structured interview is one in which questions to be asked were designated before the interview and all applicants are asked the same questions. Follow-up and probe questions are allowed, even encouraged, in highly structured interviews. Another interview design element is whether the interview is conducted by a single individual or a panel. The research literature generally supports panel interviews (e.g., multiple interviewers serving as a panel) as more reliable, valid, and useful than interviews conducted by a single individual (Huffcut, Culbertson, & Weyhrauch, 2013).

The third design issue is how the interview is scored (or not). The typical unstructured interview is not formally scored. Rather, an overall subjective "go/no go" assessment is typical. More formal methods can be used by single interviewers or a panel. For example, interviewers can rate the acceptability or quality of candidate responses on one or more dimensions. Individual judgments can be combined by some rule (averaging, for example) or discussed to reach an overall consensus in a panel interview setting. Training interviewers in the scoring system is critical.

The fourth design issue is the timing of the interview or interviews. For example, a behavioral interview with two panelists is conducted in “Step Two” of the Tigerair® pilot selection process, followed by another panel interview in “Step Three” of the selection process. As interviews are expensive (in terms of time and resources, relative to the costs of other steps in pilot selection processes), they often occur in the latter or final stages of pilot selection processes, after the numbers of applicants has been reduced.

The final design issue is how the results of the interview are used in making decisions about which applicants will, in fact, be offered employment with an operator. While there is a substantive body of empirical research on interview topics such as the psychological constructs suitable for assessment by interview, question types, structure, interviewer training, and interview scoring, there is relatively little research on how interview results are used in organizational decision-making. Interview results can be used as stand-alone “go/no go” hurdles or presented with results from other assessments for a “holistic” judgment. Another approach with scored interviews is to combine the interview with other scores in a weighted composite, where persons with composite scores at or above some pre-defined cut-off are considered for employment and persons with scores below the cut-off are rejected. Such “actuarial” approaches to decision-making in an employment context generally result in more valid and useful decisions than more subjective “clinical” or “holistic” interpretations of discrete data elements (Kuncel, Klieger, Connelly, & Ones, 2013).

There is a large and growing body of scientific research to guide development, validation, and use of interviews in personnel selection. However, it is unclear as to what degree interviews for pilot selection are informed by this body of applied psychological research.

CHAPTER 4: SCREENING IN OTHER TRANSPORTATION-RELATED OCCUPATIONS

The third chapter turns to examples of other transportation-related occupational selection programs such as air traffic controller, again based on publicly available documentation. The degree to which these programs conform to the "best practices" outlined in the first chapter is evaluated and compared to the state-of-the-art in pilot selection.

Air traffic control specialist (ATCS; FG-2152)

A majority of the research on the selection of air traffic controllers in the U.S. has been conducted at the FAA's Civil Aerospace Medical Institute (CAMI) in Oklahoma City. Additional details concerning the history of those activities can be found in Collins, Boone, & VanDeventer (1981), Broach & Manning (1998), and Sells, Dailey & Pickrel (1984). Assessment of applicant mental abilities and aptitudes was initiated following work by Brokaw (1959) who identified several aptitude tests (Arithmetic reasoning, symbolic reasoning, code translation, and an air traffic problems test) that were effective in predicting training performance. Research on controller selection at the Civil Aeromedical Institute (now called CAMI) was initiated on an experimental basis by Trites and Cobb in the 1960s (Trites, 1961, 1964; Cobb, 1962, 1964; Trites & Cobb, 1962, 1964a, 1964b). This research resulted in a decision by the Civil Service Commission (now called the Office of Personnel Management – OPM) to identify subtests from two test batteries (Differential Aptitude Test - Space Relations, Numerical Ability, and Abstract Reasoning) and the California Test Bureau's Test of Mental Maturity (Analogies, Interference, Numerical quality, and Numerical Quantity) for administration to select applicants. Results from administration of these tests proved to be effective predictors of the performance of students in the Academy air traffic training program (Cobb, 1971).

In 1962, the test battery was used to identify potential applicants from those who were not initially qualified on the basis of aviation-related experience and/or education. In 1966 the medical examination was expanded with introduction of the 16PF as a means of assessing personality attributes. A small minority of the applicants who were identified as potentially having significant emotional or mental problems were required to submit to a psychiatric examination. To further reduce attrition from basic training in 1976 a 9-week screening program was implemented at the FAA Academy. Initially, there were separate screens for those entering terminal and en route ATC facilities. Thus, at that time the controller selection process involved two major steps, the written aptitude test battery and the screening program. The ATCS screen involved training on the principles of non-radar ATC rules and requiring the trainees to utilize those rules and principles to resolve a series of laboratory simulations. Success in the screen was based on student scores on classroom tests, laboratory simulations of nonradar ATC, and a final written examination (Della Rocco, Manning, & Wing, 1990).

Dailey (1984) provides a brief description of early studies designed to assess the tasks and skills of air traffic controllers. While early research by Davis, Kerle, Silvestro, and Wallace (1960a, 1960b) was focused on assessing aspects of ATCS training, later work by Older & Cameron (1972) provided insights into the psychological dimensions of controller tasks and activities. A comprehensive job analysis of the work of ATCS in both terminal and en route environments was conducted from 1972 to 1975 by the System Development Corporations. Dailey (1984) indicates that while the investigations provided the necessary information to support enhancements to controller training, "… it was of limited value in supporting work on the basic task of selection of controllers, because of the great number and complexity of the controller skills and functions reported (p. 131)."

Efforts were initiated in 1970 to develop a new test battery for ATCS selection. The two tests that were developed as a result of this effort included the Multiplex Controller Aptitude Test (MCAT; Dailey & Pickrel, 1984a) and the Occupational Knowledge Test (OKT; Dailey & Pickrel, 1984b). The MCAT was a paper and pencil job sample that requires individuals to use table reading, visualization, assessments of traffic movement, and arithmetic reasoning to resolve a series of questions concerning the flow of air traffic on a simulated radar screen. The test was time limited. Rather than relying on an applicant's description of their ATC related experience, the OKT was developed to assess an applicant's ATC related knowledge. The final test battery was comprised of the MCAT and an abstract reasoning test (ABSR). The OKT was used to give applicants extra credit for their experience. The necessary studies were conducted to assess the overall reliability and validity of the tests. Validation of the selection procedures included use of measures of applicant performance in the Academy Screen as well as post-training performance measures that included supervisor ratings, over the shoulder ratings, and attrition. Much of that research is described in Sells, Dailey, & Pickrel (1984).

The two-stage selection process continued from 1976 to 1992. Research to assess the validity of the two-stage selection process continued throughout this time period. These assessments included evaluations of the relationship between the ATCS screen and supervisor ratings of performance (VanDeventer, 1981), between the ATCS screen score and field instructor ratings and a measure of status in en route field training (Manning, Della Rocco, & Bryant, 1989). Field training status was significantly related with both OPM and Academy scores (.30 and .44). Multiple regression analyses revealed that both measures accounted for a similar percentage of the variance in field training status. Broach and Manning (1994) called for additional research to better understand the cognitive constructs that underlie controller performance.

Consistent research outcomes by Cobb and colleagues (Cobb, Nelson, & Mathews, 1973) revealed that applicants who entered training over age 30 experienced a much higher rate of failure both in the Academy Screen and during field training. This work led Congress to establish a regulation that applicants would not be allowed to enter into ATCS training if they were over age 30. This restriction is still in place. Applicants must be between 18-30 years of age.

A review of the ATCS selection process was initiated to reduce the applicant and agency costs. The overall goal was to (a) reduce the costs, (b) maintain the validity of the ATCS selection process, and (c) support agency cultural diversity goals. This effort led to the development of a short-term approach to replacing the 9-week screen. A longer-term effort was designed to generate a one-step selection process. The interim process involved the design and development of a computer-administered test battery. The identification of components of the battery was based on a cognitive task analysis developed by Human Technology, Inc. (1991). The battery included two computer-administered information processing tests (Static Vector/Continuous Memory and Time Wall/Pattern Recognition) and a low-fidelity radar simulation of ATC vectoring and separation (Air Traffic Scenario Test (ATST) (Broach & Brecht-Clark, 1993). A predictive, criterion-related validity assessment of the test battery involved 423 newly hired ATC students who entered the 9-week screen. Prior to entry into the screen, each student received 20 practice sessions on the full battery (3 ½ days). On the following day, students received 4 sessions on the information processing tests and 6 ATST scenarios. The criterion for the evaluation was the student's final composite score from the screen. A regression analysis revealed that the computer tests scores accounted for an additional 20% of the variability in predicting the ATCS screen scores above that of the student's civil service rating (a weighted composite of OPM written test scores plus veteran's preference points; see Broach, 1998) upon hire.

A second study assessed the concurrent, criterion-related validation of the computer test battery (Weltin, Broach, Goldbach, & O'Donnell, 1992). The sample for this investigation involved 297 trainee (developmental) and full performance level (FPL) controllers. The criterion involved composite field training scores (hours of on the job training and subjective ratings of performance). When corrected for explicit and incidental restriction in range, the multiple correlation (R) was .25. The validity was equivalent to that of the 9-week ATCS screen. A follow-on analysis revealed that the validities did not vary as a function of sex or minority group status (Weltin *et al.*, 1992).

The use of the above-mentioned computer test battery was short lived. To further reduce costs to students and the FAA, the FAA embarked on developing a selection test that would serve as the equivalent of the earlier two-stage process. The instrument was designed so that individuals who entered the Academy screen had the knowledge, skills, abilities and other attributes that were equivalent to those of the existing controller workforce. Given these conditions, the Academy focus shifted from screening to training.

A multi-phase research effort was initiated in 1991 to develop a new selection instrument. In the first phase, the primary focus was execution of a job task analysis of the air traffic control position. Details concerning this effort can be found in Bobko, Nickels, Blair and Tartak (1994) and Nickels, Bobko, Blair, Sands, and Tartak (1995). A critical incident study conducted by Hedge, Borman, Hanson, Carter, and Nelson (1993) provided additional information. The resulting job/task analysis information was used in the second phase to identify (a) ATCS job performance dimensions and (b) abilities and other personal characteristics ("aptitudes") underlying performance. The second phase of the research effort began in 1996, culminating in a large sample concurrent, criterion-related validation study. Elements that were included in the overall validity study included: development of the predictor test battery; development of criterion measures; conduct of concurrent validation study; conduct of a pseudo-applicant study; analyses and validation of the predictors; and delivery of the computer-administered Air Traffic Selection and Training (AT-SAT) test battery. Details concerning these events are contained in two technical reports (Ramos, Heil & Manning, 2001a and 2001b). These reports demonstrated how the development and validation of AT-SAT complied with legal and professional guidelines. AT-SAT was comprised of 8 tests: *Applied Math*, *Dials, Angles*, *Analogy*, the *Experiences Questionnaire*, *Scan*, *Letter Factory*, and the *Air Traffic Scenarios Test*.

An important aspect of this research effort was the attention given to measuring controller job performance as reported by Borman, Hedge, & Hanson (1992) and Hedge, Borman, Hanson, Carter, and Nelson (1993). Controller performance measures used in the concurrent validation study included traditional assessments of work performance by supervisor and peers using behavioral summary scales, a medium fidelity Computer Based Performance Measure (CBPM), and a high fidelity performance measure (HFPM). The CBPM involved development of a set of air traffic scenarios using a synthetic airspace.

Controllers were allowed 60 seconds to study the flight progress strips, an airspace summary and sector map before they observed a scenario. Following each scenario, they selected from several alternatives the response that would be most effective. The responses were graded by subject matter experts regarding their overall effectiveness. Availability of the FAA Academy radar training laboratories allowed researchers to develop a high-fidelity performance measure (HFPM). Once again, a synthetic airspace was used to develop air traffic scenarios that could be used to assess important aspects of controller performance; maintaining aircraft separation, coordinating, communicating, and maintaining situation awareness. A total of 16 scenarios 60 minutes in length were developed. Controller performance was assessed using over the shoulder ratings by trained air traffic control observers. Details concerning the approach to HFPM can be found in Manning (2000).

Analyses of the validity of AT-SAT included information from the CBPM, HFPM, as well as peer and supervisor ratings. The uncorrected correlation between the weighted composite of aptitude tests and the composite of CBPM and supervisory ratings was .51 (n=1,029) (Waugh, 2j001, p. 124). When corrected or restriction in range and criterion unreliability, the correlation of the aptitude composite and composite performance measure was .76 (n=1,029 (Waugh, p. 41 and 124). In contrast, the meta-analytic estimate of combined psychomotor and cognitive ability test batteries for pilots is about .31 (.37 when corrected for dichotomization of the criterion; Martinussen, 1996). There were differences in score distributions on the final AT-SAT test battery by demographic group (specifically, females, African Americans and Hispanics scored lower than white males; see Waugh, 2001, p. 44). However, the researchers concluded that the test battery was equally valid for all groups and that mean differences in predictor scores by group were associated with comparable differences in job performance measures (Waugh, p. 47). Therefore, the research team concluded that there was strong evidence that the AT-SAT test battery was fair for protected groups (Waugh, p. 47). Subsequent investigators (Manning & Heil, 2001) also determined the relationship between AT-SAT and archival criterion measures (OPM Test Battery, Academy Nonradar Screening Program, The Pre-Training Screen, Radar Training as well as other archival tests. Significant correlations were found between AT-SAT and the various training measures as well as between various subtests and experimental tests that had been administered upon entry into the Academy. This effort represented the first complete approach to the development and validation of a U.S. air traffic controller selection test battery where the job task analysis was used to link controller performance with the development of the battery as well as development of both medium fidelity (CBPM) and high fidelity (HFPM) criterion performance measures.

In accordance with best practices, work continued on assessing the effectiveness of AT-SAT along with other potential measures including the use of biographical data (Broach, Byrne, Manning, Pierce, McCauley, & Bleckley, 2013; Broach, 2012; Pierce, Broach, Byrne, & Bleckley, 2014). In addition, research has also been directed toward the potential implications of the introduction of new equipment and procedures associated with the Next Generation Air Transportation System (“NextGEN”; Broach, 2013).

As evidenced by this record, substantial attention has been paid over the years to cognitive abilities in the ATCS occupation. Less attention has been given to personality. Convey (1984) provides a historical overview of the implementation and use of Cattell’s16PF as part of the controller selection process. From 1965 to 1970 the 16PF was routinely administered to applicants as part of the medical examination. The 16PF was designed to identify applicants who needed further psychiatric/psychological evaluation to determine their fitness to become ATCSs. While initially using a cut score based on a second-order anxiety factor, concerns about the about the validity and primary structure of the instrument led Dailey to develop a specialized scoring approach based on what he called iterative distillation. The revised scoring system remained in place until the 16PF was replaced by the Multiphasic Personality Inventory (MMPI) as part of the pre-employment medical examination in 2008 (see King, Schroeder, Manning, Retzlaff, & Williams, 2008).

Additional investigations were initiated to assess the effectiveness of other measures in predicting success in the Academy Screen and during field training. In general, those efforts yielded limited success. The investigations included use of the 16PF (Smith, 1971; Stark, 1994; Broach, 2006), Myers Briggs Type Indicator (Schroeder & Dollar, 1997; Dollar & Schroeder, 2004), the State-Trait Personality Indicator (Collins, Schroeder, & Nye, 1991), and NEO-PI (Schroeder, Broach, & Young, 1993). Individuals with higher scores on the trait anxiety measure of the STPI were generally less successful in both the Academy Screen and field training. With the 16PF the scale with the highest correlation with success was Factor B, which provides a rough measure of general intelligence.

Railroad dispatcher

Railroad or train dispatchers play an important role in the U.S. transportation system. Like air traffic controllers, they are responsible for the safe and efficient movement of railroad traffic within a specific geographic region. However, while most air traffic controllers are employed the U.S. federal government, train dispatchers are employed in the private sector by railroad operating companies such as CSX, Inc., BNSF, Inc., and many smaller regional and local operators.

The Federal Railroad Administration conducted research on the selection of train dispatchers in the mid- and late-1990s, culminating in the report “Selection of Railroad Dispatcher Candidates” (Gertler, 2003). Much like this report, the FRA report outlined a best practices framework for personnel assessment and selection programs. However, unlike this report, the FRA study team also conducted a formal analysis of the dispatcher job using the Position Analysis Questionnaire (PAQ®; McPhail, Jeanneret, McCormick, & Mecham, 1998), a widely used structured job analysis technique to identify attributes to be assessed in a dispatcher selection process. The study team then evaluated the dispatcher selection programs for seven operators relative to the best practices framework and attributes identified by the PAQ® analysis as important to successful performance of the train dispatcher job. The FRA team found that larger operators had more formal selection programs, sometimes involving formal psychometric ability testing. Interviews were very common regardless of operator size.

The PAQ® job analysis indicated specific abilities were relevant and important to the train dispatcher job. Simple reaction time was the only ability from the psychomotor domain identified by the PAQ® as important to train dispatching. There were no significant physical ability requirements, such as strength or stamina. The most important sensory ability was auditory acuity (hearing acuity). Perceptual abilities identified as important to the train dispatcher job were perceptual speed, selective attention, and time-sharing (e.g., multi-tasking). Important cognitive abilities in train dispatching were short-term memory, long-term memory, overall intelligence, and convergent thinking.

The study team then made general recommendations about how to develop dispatcher selection programs. General recommendations were made about ability testing based on the results of the PAQ® job analysis. Interestingly, the FRA study team specifically examined FAA air traffic control specialist selection processes and research, just as we have done in this report. However, it is unclear as to what happened after the report was issued. No follow-up studies were found in either the Transportation Research Board library or search via Google Scholar.

Comparison of selection of pilots with selection in other transportation occupations

The purpose of this chapter was to compare selection of pilots to selection in other transportation occupations. There are several points to be considered. First, there is an extensive research record for pilot selection. The only comparable transportation occupation with such a research record is the ATCS occupation in the FAA. The track record for research on selection in other occupations, such as railroad dispatcher, pipeline dispatch, truck driver, locomotive engineer, and commercial bus driver is elusive. In part, this is because selection processes for these other transportation occupations is decentralized across multiple private entities (corporations), and in some cases, across states (for example, state standards for commercial drivers' licenses). Second, despite the volume of pilot selection research, much of the information publicly available is at a very high level with little supporting detail, with the exception of U.S. military pilot selection research. Third, substantial work has been invested in the measurement of air traffic controller job performance. Less attention has been given in the pilot community to actual pilot job performance. If the goal is to predict not just who succeeds (or fails) in pilot training, greater attention to developing measures of pilot job performance in the cockpit is needed (Broach, 2014). Finally, it is clear from the research record that cognitive abilities are the prime determinants of performance in training and on the job (to the extent that job performance is measured) for both pilots and air traffic controllers.

CHAPTER 5: SUMMARY, RECOMMENDATIONS, AND RESEARCH QUESTIONS

In this final chapter, we summarize our evaluations of pilot selection processes relative to the "best practices" framework, make several recommendations, and suggest future research.

Summary of best practices & issues

Job analysis. Relatively few of the test batteries in use by either flight schools or operators are based on a recent, selection-oriented job/task analysis. This is a significant conceptual and legal vulnerability. On one hand, job/task analysis provides the conceptual basis for a selection system through the identification of the knowledge, skills, abilities, and other personal characteristics underlying successful performance in training and on the job. Without a job/task analysis to guide predictor selection and frame performance assessment, an employer is groping in the dark on the basis of anecdote, intuition, and past experience. On the other hand, a selection procedure challenged in U.S. federal courts as having adverse impact on one or protected classes is highly unlikely to survive that review without a supporting job/task analysis.

Predictors. On one hand, the predictors in most of the psychometric test batteries are reasonable and relate to the "*g*" factor identified by Ree and Carretta as the primary psychological construct underlying pilot training outcomes and performance. On the other hand, interviews are both widely used and poorly documented. The reliability, validity and overall utility of interviews in civilian pilot selection is not well established by empirical evidence.

Criteria. The primary criterion for validation of the test batteries is success in *ab initio* and other employer-specific flight training and during the initial operating experience period. While useful, success in training merely indicates that the pilot can perform flying tasks at some minimally acceptable level. Clear operational definitions and measures of terms such as "basic airmanship skills" found in U.S. law and regulations are desperately needed to improve the state-of-the-art in pilot selection.

Fairness. No data were available on comparative selection (pass) rates by demographic group on the test batteries or interviews. This represents a significant vulnerability in the U.S., as the lack of such data can be used by a fact finder, under some circumstances, as the basis for an adverse finding against an employer in case of a legal challenge in federal proceedings (see. 29 C.F.R. § 1607.3D).

Utility. Utility refers to the pay-off that can be expected from a selection test. Example measures of utility include increases in training success rate and reductions in the cost of training. Little to no information is available on the utility of pilot selection test batteries and interviews. The one exception is the DLR pilot test battery, with several formal papers on the financial utility of valid pilot selection.

Reporting. Reporting is thin on several of the test batteries, consisting largely of marketing-oriented documents. However, data on PILAPT®, COMPASS, ADPAT™, and the DLR pilot selection process have been presented at technical conferences and in scientific publications. Reporting is even thinner for interviews as elements of pilot selection processes. Limited technical information is available for the pilot selection process for one U.S. mainline air carrier.

Recommendations

Future research could potentially illuminate challenges and lead to improved selection procedures for pilots in the U.S. (and international) aviation industry. The most critical issue is the operational definition and measurement of pilot performance at different stages of training and in an operational environment. Training success overall is a necessary but insufficient criterion. Finer-grained measures of technical flying performance and CRM behaviors in the cockpit at each stage of training and in revenue service would provide invaluable feedback for an evidence-based training process. Moreover, more detailed performance measures could provide a basis for identifying trends in pilot performance when examined across time, pilots, and operators. Such pilot performance data, going beyond simple flight check data (satisfactory, unsatisfactory), could be accumulated in a Pilot Records Database, for example. Such measures might be mined (analyzed) at some future point for identifying performance dimensions in lieu of flight hours *per se* as a basis for pilot certification.

The second critical issue is the documentation of pilot selection systems, particularly in the US. The data on pilot selection in the U.S. appears to be hidden largely behind corporate firewalls, in part because such processes are part and parcel of corporate human resource management strategies and are viewed as a source of competitive advantage in the so-called "war for talent." As a result, the scientific knowledge base for civilian pilot selection is far less developed than the knowledge base for military pilot selection. The consequence, as noted by IATA in 2012, is selection programs without a strong scientific basis. We suggest building a clearinghouse of pilot selection data modeled on the Aviation Safety Information and Analysis Sharing (ASIAS) and the Flight Operations Quality Assurance (FOQA) programs. Information and data on pilot selection processes might be provided by Part 121 operators and pilot training programs under a confidentiality agreement. Analyses of such de-identified data might then be provided back to industry partners to inform continuing development and enhancement of their pilot selection processes.

Future research questions

Overall, there appear to be two basic models in civilian pilot selection. The Lufthansa model is stringent selection on the front-end with less attrition in training. The alternate model is cursory selection on the front-end with more attrition in training. Indeed, at least in the U.S., the pilot training process itself is the primary selection mechanism for determining if a pilot can safely operate in the NAS. The progression from private, instrument, commercial, and ATP provides the regulator with repeated opportunities to evaluate a pilot's ability to safely operate aircraft to increasingly tighter standards. This suggests the following research questions.

Do the current practical test standards measure competencies required in the respective types of operations (e.g. Part 91, 135, 121 and the subsets thereof)?

As Part 135, Part 121, and in many instances Part 91 operations gravitate toward tasks that are more heavily dependent on system setup and monitoring, are current practical test standards adequately measuring the behaviors required to safely fly the line?

Are there key differences between flight operations that should be explored in future task analyses (e.g. Part 91 corporate pilots, Part 135, Part 121)? If differences exist, do these auger for modified selection criteria?

What is the impact of increasingly sophisticated flight control and management technology on pilot performance and thus on the knowledge, skills, abilities, and other personal characteristics required of pilots? How do we represent and measure human performance in a highly automated cockpit, where in some sense the computer does the flying?

These are just a few of the questions that might be generated when considering future research on pilot selection processes. Flight technologies are likely to continue to advance and evolve rapidly. Following the best practices described in this report – analyze the pilot's job, measure job performance, select and test predictors based on the job analysis, validate predictors against job performance, evaluate fairness and utility, and finally, document and report the analyses to expand the knowledge base and continually improve pilot selection processes – will be helpful to organizations adapting to these technological advances and ensuring that persons with "the right stuff" (whatever that might look like) are in commercial cockpits in the future.

REFERENCES

Air Line Pilot's Association, International. (2009). *Producing a professional airline pilot: Candidate screening, hiring, training and mentoring.* Retrieved from (https://public.alpa.org/portals/alpa/pressroom/inthecockpit/ProducingProfessionalPilot_9-2009.pdf)

Air Line Pilot's Association, International. (2015, June 25). *Finding and keeping the best and brightest in the cockpit.* Retrieved from (https://www.alpa.org/news-and-events/Blog/2015/06/25/keeping-the-best-and-brightest)

Air Line Pilot's Association, International. (2016, March 18). *Keeping the pilot profession strong.* Retrieved from (https://www.alpa.org/news-and-events/Blog/2016/03/18/keeping-the-pilot-profession-strong)

Albermale Paper Co. v. Moody, 422 U.S. 405 (1975).

Albert Consulting. (2015). *Pilot selection.* Retrieved from (http://albertconsulting.se/pilot.htm)

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for educational and psychological testing.* Washington, DC: American Psychological Association.

Arendasy, M., Sommer, M., & Hergovich, A. (2007). Statistical judgment formation in personnel selection: A study in military aviation psychology. *Military Psychology*, *19*, 119-136.

Arthur, W. Jr., & Day, E.A. (2011). Assessment centers. In S. Zedeck (Ed.), *APA Handbook of Industrial and Organizational Psychology Volume 2: Selecting and developing members for the organization* (pp. 205-235). Washington, DC: American Psychological Association.

Bartram, D., & Baxter, P. (1996). Validation of the Cathay Pacific Airways pilot selection program. *International Journal of Aviation Psychology*, *6*, 149-169.

Biddle, D.A. (2010). Should employers rely on local validation studies or validity generalization (VG) to support the use of employment tests in Title VII situations? *Public Personnel Management*, *39*(4), 307-326.

Bjerke, E., Smith, G., Smith, .M.J, Christensen, C., Carney, T., Craig, P., & Niemczyk, M. (2016). Pilot Source Study 2015: US regional airline pilot hiring background characteristic changes consequent to Public Law 111-216 and the FAA First Officer Qualifications Rule. *Journal of Aviation Technology and Engineering*, *5*(2), 1-12. Retrieved from (https://docs.lib.purdue.edu/jate/vol5/iss2/1/)

Blair, J., & Freye, J. (2012, March). *Flight training capacity in the context of recent legislation: An examination of the impacts of reduced training capacity, and the declining rates of airmen certification.* Portage, MI: National Association of Flight Instructors. Retrieved from (http://www.nafinet.org/File/Flight%20Training%20Capacity%20in%20the%20Context%20of%20Recent%20Legislation%20-%20by%20Jason%20Blair%20and%20Jonathon%20Freye.pdf)

Bobko, P., Nickels, B.J., Blair, M.D., & Tartak, E.L. (1994). *Preliminary internal report on the current status of the SACHA model and task interconnections: Vol. 1.* Bethesda, MD: The University Research Corporation (URC).

Boeing. (2014). *2014 Pilot & technician outlook.* Chicago, IL: Boeing Flight Services.

Borman, W.C., Hedge, J.W., & Hanson, M.A. (1992). *Criterion development in the SACHA project: Toward accurate measurement of air traffic control specialist performance* (Institute Report #222). Minneapolis, MN: Personnel Decisions Research Institutes.

Boudreau, R. (July 13, 2014). IFALPA flags up concerns over MPL training. *AINonline.* Retrieved from (https://www.ainonline.com/aviation-news/air-transport/2014-07-13/ifalpa-flags-concerns-over-mpl-training)

Broach, D. (1998). Air traffic control specialist aptitude testing, 1981-1992. In D. Broach (Ed.), *Recovery of the FAA air traffic control specialist workforce, 1981-1992* (Report No. DOT/FAA/AM-98/23). (pp. 7-16). Washington, DC: Federal Aviation Administration Office of Aviation Medicine. Retrieved from (http://www.faa.gov/data_research/research/med_humanfacs/oamtechreports/1990s/media/AM98-23.pdf)

Broach, D. (2006, June). *Validity of personality test scores at hire for prediction of subsequent claims for work-related "mental, nervous or emotional disorders" among air traffic controllers.* Poster presented at the 26th International Congress for Applied Psychology, Athens, Greece.

Broach, D. (2012). *Incremental validity of biographical data in the prediction of en route air traffic controller technical knowledge and skills.* (Report No. DOT/FAA/AM-12/8). Washington, DC: Federal Aviation Administration Office of Aerospace Medicine. Retrieved from (http://www.faa.gov/data_research/research/med_humanfacs/oamtechreports/2010s/media/201208.pdf)

Broach, D. (2013). *Selection of the next generation of Air Traffic Control Specialists: Aptitude requirements for the air traffic control tower cab in 2018.* (Report No. DOT/FAA/AM-13/5). Washington, DC: Federal Aviation Administration Office of Aerospace Medicine. Retrieved from (http://www.faa.gov/data_research/research/med_humanfacs/oamtechreports/2010s/media/201305.pdf)

Broach, D. (2014). Commentary on the article by Weismuller and Damos: Statistical models and the structure of selection processes in pilot selection systems. *International Journal of Aviation Psychology*, *24*, 119-122.

Broach, D., & Brecht-Clark, J. (1993). Validation of the Federal Aviation Administration air traffic control specialist Pre-Training Screen. *Air Traffic Control Quarterly*, *1*, 115 - 133.

Broach, D., Byrne, C., Manning, C., Pierce, L., McCauley, D., & Bleckley, M.K. (2013). *The validity of the Air Traffic Selection and Training (AT-SAT) test battery in operational use.* (Report No. DOT/FAA/AM-13/3). Washington, DC: Federal Aviation Administration Office of Aerospace Medicine. Retrieved from (http://www.faa.gov/data_research/research/med_humanfacs/oamtechreports/2010s/media/201303.pdf)

Broach, D., & Manning, C. A. (1994). *Validity of the air traffic control specialist (ATCS) Nonradar Screen as a predictor of performance in radar-based air traffic control training.* (Report No. DOT/FAA/AM-94/9). Washington, DC: Federal Aviation Administration Office of Aviation Medicine. Retrieved from (http://www.faa.gov/data_research/research/med_humanfacs/oamtechreports/1990s/media/AM94-04.pdf)

Broach, D. & Manning, C. (1998). Issues in the selection of air traffic controllers. In M.W. Smolensky & E.S. Stein (Eds.), *Human factors in air traffic control* (pp. 237-271). San Diego, CA: Academic Press.

Brokaw, L.D. (1959, April). *School and job validation of selection measures for air traffic control training.* (WADC-TN-59-39). Lackland AFB, TX: U.S. Air Force Personnel Laboratory.

Burger, K.-H., Neb, H., & Hörmann, H.-J. (2002). *Basic performance of flight crew: A new concept of competence based markers for defining pilots performance profile.* Proceedings of the 25th symposium of the European Association for Aviation Psychology, Zegrze, Poland.

Burger, K.-H., Neg, H., & Hörmann, H.-J. (2003). Lufthansa's new basic performance of flight crew concept: A competence based marker system for defining pilots' performance profile. *Proceedings of the 12th International Symposium on Aviation Psychology* (Vol. 1, pp. 172-175). Dayton, OH: Wright State University.

Burke, E., Hobson, C., & Linsky, C. (1997). Large sample validations of three general predictors of pilot training success. *International Journal of Aviation Psychology*, *7*(3), 225-234.

Burke, E., Kitching, A., & Vaslsler, C. (1995). Computer-based assessment and construction of valid aviation selection tests. In N. Johnston, R. Fuller, & N. McDonald (Eds.). *Aviation psychology: Training and selection* (pp. 282-287). Cambridge, United Kingdom: Avebury Aviation.

Burke, E. F., Kitching, A. & Valsler, C. (1997). The Pilot Aptitude Tester (PILAPT): On the development and validation of a new computer-based test battery for selecting pilots. *Proceedings of the Ninth International Symposium on Aviation Psychology* (pp. 1286-1291). Columbus, OH: The Ohio State University.

Burke, E. F., Kitching, A., & Valsler, C. (2001). Pilot selection: Getting more bang for the buck. In Hayward, B. H., & Lowe, A. R. (Eds.) Aviation Resource Management: *Proceedings of the Fourth Australian Aviation Psychology Symposium* (pp. 9-20). Burlington, VT: Ashgate.

Campion, M.A., Campion, J.E., & Hudson, J.P. Jr. (1994). Structured interviewing: A note on incremental validity and alternative question types. *Journal of Applied Psychology*, *79*, 998-1002.

Cascio, W.F. (1999). *Costing human resources: The financial impact of behavior in organizations* (4th Ed.). Cincinnati, OH: South-Western College Publications.

Chidester, T.R., Helmreich, R.L., Gregorich, S.E., & Geis, C. E. (1991). Pilot personality and crew coordination: Implications for training and selection. *International Journal of Aviation Psychology*, *1*(1), 25-44.

Chidester, T.R., Kanki, B.G., Foushee, H.C., Dickinson, C.L., & Bowles, S.V. (1990). *Personality factors in flight operations: I. Leader characteristics and crew performance in full-mission air transport simulation.* (NASA Technical Memorandum No. 102259). Moffett Field, CA: NASA Ames Research Center. Retrieved from (https://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/19900014054.pdf)

Cobb, B.B. (1962). Problems in air traffic management: II. Prediction of success in air traffic controller school. *Aerospace Medicine*, *33*, 702-713.

Cobb, B.B. (1964). Problems in air traffic management: V. Identification and potential of aptitude test measures for selection of tower air traffic controller trainees. *Aerospace Medicine*, *35*, 1019-1027.

Cobb, B.B. (1971). *Air traffic aptitude test measures of military and FAA controller trainees.* (Report No. DOT/FAA/AM-71/40). Washington, DC: Federal Aviation Administration Office of Aviation Medicine. Retrieved from (http://www.faa.gov/data_research/research/med_humanfacs/oamtechreports/1970s/media/AM71-40.pdf)

Cobb, B. B., Nelson, P. L., & Mathews, J. J. (1973). *The relationships of age and ATC experience to job performance ratings of terminal area traffic controllers.* (Report No. DOT/FAA/AM-73/7). Washington, D.C.: Federal Aviation Administration Office of Aviation Medicine. Retrieved from (http://www.faa.gov/data_research/research/med_humanfacs/oamtechreports/1970s/media/AM73-07.pdf)

Collins, W.E., Boone, J.O., & VanDeventer, A.D. (1981). The selection of air traffic control specialists: History and review of contributions by the Civil Aeromedical Institute, 1960-1980. *Aviation, Space, and Environmental Medicine*, *52*, 217-240.

Collins, W. E., Schroeder, D.J., & Nye, L.G. (1991). Relationships of anxiety scores to Academy and field training performance of air traffic control specialists. *Aviation, Space, and Environmental Medicine*, *62*, 236-240.

Collup, J. W. (2007). *Forecasting Demand for Civilian Pilots: A Cost Savings Approach to Managing Air Force Pilot Resources* (Report No. AFIT/GCA/ENV/07-M4). Wright-Patterson AFB, OH: The U.S. Air Force Air University, Institute of Technology. Retrieved from (http://www.dtic.mil/get-tr-doc/pdf?AD=ADA467529)

Convey, J.J. (1984). Personality assessment of ATC applicants. In S.B. Sells, J.T. Dailey, & E.W. Pickrel, E.W. (Eds.), *Selection of air traffic controllers* (pp. 323-351). (Report No. DOT/FAA/AM-84/2). Washington, DC: Federal Aviation Administration Office of Aviation Medicine. Retrieved from (http://www.faa.gov/data_research/research/med_humanfacs/oamtechreports/1980s/media/AM84-02.pdf)

Cooper, G.E., White, M.D., & Lauber, J.K. (Eds.). (1979). *Resource management on the flightdeck.* (NASA Conference Publication 2120). Moffett Field, CA: NASA Ames Research Center. Retrieved from (https://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/19800013796.pdf)

Croft, J. (2015a, February 3). U.S. carriers face shrinking pool of pilots. *Aviation Week & Space Technology*, p. 62.

Croft, J. (2015b, February 3). As airline pilot shortage looms, industry examines its options. *Aviation Week & Space Technology*, p. 68.

Dahlstrom, N., Laursen, J., & Bergstrom, J. (2008). *Crew Resource Management, threat and error management, and assessment of CRM skills. Current situation and development of knowledge, methods and practices. Report for the Swedish Civil Aviation Authority.* Lund University School of Aviation (LUSA). Retrieved from (http://lup.lub.lu.se/luur/download?func=downloadFile&recordOId=1981926&fileOId=1981944)

Dailey, J.T. (1984). Characteristics of the air traffic controller. In Sells, S.B., Dailey, J.T., & Pickrel, E.W. (Eds.). *Selection of air traffic controllers* (pp. 128-141). (Report No. DOT/FAA/AM-84/2). Washington, DC: Federal Aviation Administration Office of Aviation Medicine. Retrieved from

(http://www.faa.gov/data_research/research/med_humanfacs/oamtechreports/1980s/media/AM84-02.pdf)

Dailey, J.T., & Pickrel, E.W. (1984a). Development of the Multiplex Controller Aptitude Test. In S.B. Sells, J.T. Dailey, & E.W. Pickrel (Eds.), *Selection of air traffic controllers* (pp. 281-298). (Report No. DOT/FAA/AM-84/2). Washington, DC: Federal Aviation Administration Office of Aviation Medicine. Retrieved from (http://www.faa.gov/data_research/research/med_humanfacs/oamtechreports/1980s/media/AM84-02.pdf)

Dailey, J.T., & Pickrel, E.W. (1984b). Development of the air traffic controller Occupational Knowledge Test. In S.B. Sells, J.T. Dailey, & E.W. Pickrel (Eds.), *Selection of air traffic controllers* (pp. 299-321). (Report No. DOT/FAA/AM-84/2). Washington, DC: Federal Aviation Administration Office of Aviation Medicine. Retrieved from (http://www.faa.gov/data_research/research/med_humanfacs/oamtechreports/1980s/media/AM84-02.pdf)

Damitz, M., Manzey, D., Kleinmann, M., & Severin, K. (2003). Assessment center for pilot selection: Construct and criterion validity and the impact of assessor type. *Applied Psychology: An International Review*, *52*, 193-212.

Damos, D.L. (1996). Pilot selection batteries: Shortcomings and perspectives. *International Journal of Aviation Psychology*, *6*, 199-209.

Damos, D.L. (2003, February). Pilot selection systems help predict performance. *Flight Safety Digest*, 1-13.

Damos, D. (2011). The right stuff. *Civil Aviation Training*, 5, pp. 20, 21, 23, 24.

Damos, D.L. (2014). Editor's preface to the special issue on pilot selection. *International Journal of Aviation Psychology*, *24*, 1-5.

Damos, D. L., & Koonce, J. M. (1997). Methodological and analytical concerns on the pilot selection research of Park and Lee (1992). *Human Factors*, *39*, 9-13.

Davis, C. G., Kerle, R. H., Silvestro, A. W., & Wallace, W. H. (March 15, 1960a). *The air traffic control training program as viewed by training supervisors.* (Technical report No. 33, Project O under FAA contract FAA/BRD-40). Philadelphia, PA: Courtney and Company.

Davis, C. G., Kerle, R., Silvestro, A. W., & Wallace, W. H. (1960b). *Identification of training requirements in air traffic control.* (Technical report 36, Project O under FAA contract FAA/BRD-40). Philadelphia, PA: Courtney & Co. (NTIS AD 260 533).

De Corte, W., Lievens, F., & Sackett, P. R. (2007). Combining predictors to achieve optimal trade-offs between selection quality and adverse impact. *Journal of Applied Psychology*, 1380–1393. doi:10.1037/0021-9010.92.5.1380

Della Rocco, P.S., Manning, C.A., & Wing, H. (1990). *Selection of air traffic controllers for automated systems: Applications from current research.* (Report No. DOT/FAA/AM-90/13). Washington, DC: Federal Aviation Administration. Retrieved from

(http://www.faa.gov/data_research/research/med_humanfacs/oamtechreports/1990s/media/AM90-13.pdf)

Deutsche Forschungsanstalt fuer Luft- und Raumfahrt e.V. (DLR). (2015). *Staff selection.* Retrieved from (http://www.dlr.de/me/en/desktopdefault.aspx/tabid-5070/8492_read-14798/)

Dipboye, R.L. (1994). Structured and unstructured interviews: Beyond the job-fit model. *Research in Personnel and Human Resource Management*, *12*, 79-123.

Doat, B. (1995). Need of new development in Air France selection. In N. Johnston, R. Fuller, & N. McDonald (Eds.), *Aviation psychology: Training and selection, Proceedings of the 21st Conference of the European Association for Aviation Psychology (EAAP)* (Vol. 2, pp. 182-187). Aldershot, United Kingdom: Avebury Aviation.

Dollar, C.S., & Schroeder, D.J. (2004). *A longitudinal study of Meyers-Briggs personality types in air traffic controllers.* (Report No. DOT/FAA/AM-04/21). Washington, DC: Federal Aviation Administration Office of Aviation Medicine. Retrieved from http://www.faa.gov/data_research/research/med_humanfacs/oamtechreports/2000s/media/0421.pdf

Duggar, J.W., Smith, B.J., & Harrison, J. (2011, July). International supply and demand for U.S. training commercial pilots. *Journal of Aviation Management and Education*, *1*. Retrieved from (http://www.aabri.com/manuscripts/09349.pdf)

European Pilot Selection & Training. (2015). *Computerized Pilot Aptitude Screening System* (COMPASS). Retrieved from (http://epst.com/en/saa-first-officer-training/compass-test#.Vg1BNfm3t8F)

Federal Aviation Administration. (2016). *FAA aerospace forecast: Fiscal years 2016-2036*. Retrieved from (http://www.faa.gov/data_research/aviation/aerospace_forecasts/media/FY2016-36_FAA_Aerospace_Forecast.pdf)

Flin, R., Goeters, K., Hörmann, H.-J., & Martin, L. (1998). *A generic structure of non-technical skills for training and assessment.* Paper presented at the 23rd Conference of the European Association for Aviation Psychology, Vienna, Austria.

Flin, R., & Martin, L. (2001). Behavioural markers for crew resource management: A review of current practice. *International Journal of Aviation Psychology*, *11*(1), 95-118.

Flin, R., Martin, L., Goeters, K., Hörmann, H., Amalberti, R., Valot, C., & Nijhuis, H. (2003). Development of the NOTECHS (nontechnical skills) system for assessing pilots' skills. *Human Factors and Aerospace Safety*, *3*, 97-119.

Franz, T.M., Prince, C., Cannon-Bowers, J.A.., & Salas, E. (1990). The identification of aircrew coordination skills. *Proceedings of the 12th Symposium on Psychology in the Department of Defense* (pp. 97-101). Retrieved from (http://www.dtic.mil/get-tr-doc/pdf?AD=ADA240113)

Frey, B.F., Thomas, M., Walton, A.J., & Wheeler, A. (2001). WOMBAT as an example of situational awareness testing in pilot selection: An argument for the alignment of selection training, and performance. *Proceedings of the 11th International Symposium on Aviation Psychology* (pp. 1-6). Columbus, OH: The Ohio State University.

Gertler, J. (2003, March). *Selection of railroad dispatcher candidates.* (Report No. DOT/FRA/ORD-03/09). Washington, DC: Federal Railroad Administration Office of Research and Development.

Gessellschaft für Angewandte Psychogische Forschung [Society for Applied Psychological Research] (GAPF) (2015). Retrieved from (http://www.gapf.de/aero/aeroEnglish.html)

Goeters, K.-M. (2004). Non-Technical skills assessment in pilot training: Theory and practice of the NOTECHS method. In K.-M. Goeters (Ed.), *Aviation Psychology: Practice and Research* (pp. 241-243). Burlington, VT: Ashgate.

Goeters, K.-M., & Mashcke, P. (2004). Cost benefit analysis of pilot selection: The economic value of psychological testing. In K.-M. Goeters (Ed.), *Aviation psychology: Practice and research* (pp. 203-208). Burlington, VT: Ashgate.

Goeters, K.-M., Maschke, P., & Eiβfeldt, H. (2004). Ability requirements in core aviation professions: Job analyses of airline pilots and air traffic controllers. In K.-M. Goeters (Ed.), *Aviation psychology: Practice and research* (pp. 99-119). Burlington, VT: Ashgate.

Gontar, P., & Hörmann, H. J. (2014). Flight crew performance and CRM ratings based on three different perceptions. In *Aviation Psychology: facilitating change (s): Proceedings of the 31st EAAP Conference* (pp. 310-316). Retrieved from (https://www.researchgate.net/profile/Patrick_Gontar/publication/272027651_Flight_Crew_Performance_and_CRM_Ratings_Based_on_Three_Different_Perceptions/links/5509abce0cf26198a639195e.pdf)

Guion, R.M. (1998). *Assessment, measurement, and prediction for personnel decisions.* Mahwah, NJ: Lawrence Erlbaum Associates.

Guttman, A., & Dunleavy, E.M. (2012). Documenting work analysis projects: A review of strategy and legal defensibility for personnel selection. In M.A. Wilson, W. Bennett, Jr., S.G. Gibson, & G.M. Alliger (Eds.), *The handbook of work analysis: Methods, systems, applications and science of work measurement in organizations* (pp. 139-168). New York, NY: Routledge.

Hedge, J.W., Borman, W.C., Hanson, M.A., Carter, G.W., & Nelson, L.C. (1993). *Progress toward development of ATCS performance criterion measures.* (Institute Report No. 235). Minneapolis, MN: Personnel Decisions Research Institute.

Hedge, J.W., Bruskiewicz, K.T., Borman, W.C., Hanson, M.A., Logan, K.K., & Siem, R. (2000). Selecting pilots with crew resource management skills. *International Journal of Aviation Psychology*, *10*, 377-392.

Helmreich, R.L., & Foushee, H.C. (1993). Why crew resource management? Empirical and theoretical bases of human factors training in aviation. In E.L. Wiener, B.G. Kanki, & R.L. Helmreich (Eds.), *Cockpit Resource Management* (pp. 3-45). San Diego, CA: Academic.

Helmreich, R.I., Merritt, A.C., & Wilhelm, J.A. (1999). The evolution of crew resource management training in commercial aviation. *International Journal of Aviation Psychology*, *9*(1), 19-32.

Helmreich, R.I., Wilhelm, J.A., Klinect, J.R., & Merritt, A.C. (2001). Culture, error, and crew resource management. In E. Salas, C.A. Bowers, & E. Edens (Eds.), *Improving teamwork in organizations: Application of resource management training* (pp. 305-334). Hillsdale, NJ: Erlbaum.

Higgins, J., Lovelace, K., Bjerke, E., Lounsberry, N., Lutte, R., Friedenzohn, D., ... & Craig, P. (2013). *An Investigation of the United States Airline Pilot Labor Supply*. Auburn, AL: Aviation Accreditation Board International. Retrieved from (http://www.aabi.aero/AirlinePilotLaborSupply1.pdf)

Hilkevitch, J. (August 9, 2013). Experts say stricter FAA rules for pilots too costly, won't improve safety. *Chicago Tribune*. Retrieved from (http://articles.chicagotribune.com/2013-08-09/news/ct-met-airline-pilot-training-20130809_1_regional-carriers-pilots-flight-time)

Hoffmann, C.C., Spetz, S.H., & Hoffmann, A.K. (2011). Delta air lines pilot selection: 10 years of validation results. *Proceedings of the 16th International Symposium of Aviation Psychology* (pp. 292-297). Columbus, OH: The Ohio State University.

Hörmann, H.-J. (2004). JAR-TEL results: Testing the cultural robustness of the NOTECHS method. In K.-M. Goeters, (Ed.), *Aviation Psychology: Practice and Research* (pp. 273-285). Burlington, VT: Ashgate.

Hörmann, H.-J., Burger, K.H., & Neb, H. (2003). Evaluation of pilot's interpersonal skills. Usability of an integrated proficiency reporting from within Lufthansa. *Proceedings of the 12th International Symposium on Aviation Psychology* (pp. 565-568). Columbus, OH: The Ohio State University.

Hörmann, H.-J., & Goerke, P. (2014). Assessment of social competence for pilot selection. *International Journal of Aviation Psychology*, *24*(1), 6-28.

Hörmann, H.-J., & Luo, X.L. (1999). Development and validation of selection methods for Chinese student pilots. *Proceedings of the 10th International Symposium on Aviation Psychology* (pp. 571-576). Columbus, OH: The Ohio State University.

Hörmann, H.-J., Manzey, D., Maschke, P., & Pecena, Y. (1997). Behavior-oriented assessment of interpersonal skills in pilot selection: Concepts, methods, and empirical findings. *Proceedings of the 9th International Symposium on Aviation Psychology* (pp. 1110-1115). Columbus, OH: The Ohio State University.

Hörmann, H.-J., & Maschke, P. (1996). On the relation between personality and job performance of airline pilots. *International Journal of Aviation Psychology*, *6*, 171-178.

Huffcut, A.I. (2011). An empirical review of the employment interview construct literature. *International Journal on Selection and Assessment*, *19*, 62-81.

Huffcutt, A.I. & Culbertson, S.S. (2010). Interviews. In S. Zedeck (Ed.), *APA Handbook of Industrial and Organizational Psychology* (Volume 2) (pp. 185-203). Washington, DC: American Psychological Association.

Huffcut, A.I., Culbertson, S., & Weyhrauch, W.S. (2013). Employment interview reliability: New meta-analytic estimates by structure and format. *International Journal of Selection and Assessment*, *21*, 264-276.

Human Technology, Inc. (1991, June). *Cognitive task analysis of en route air traffic control: Model extension and validation. Executive summary*. (Contract deliverable under OPM-87-9041 to the Federal Aviation Administration Office of Higher Education and Training (AHT)). McLean, VA: Author.

Hunter, D.R., & Burke, E.F. (1994). Predicting aircraft pilot-training success: A meta-analysis of published research. *International Journal of Aviation Psychology*, *4*, 297-313.

Hunter, D.R., & Burke, E.F. (1995). *Handbook of pilot selection.* Burlington, VT: Ashgate.

International Civil Aviation Organization (ICAO). (2006, November). *Procedures for Air Navigation Services - Training* (PANS-TRG). (Doc. No. 9868). Montreal, Canada: Author.

International Air Transport Association (IATA). (2012). *Guidance material and best practices for pilot aptitude testing* (2nd Ed.). Retrieved from (http://www.iata.org/publications/Documents/pilot-aptitude-testing-guide.pdf)

Kay, G.G. (1995). *CogScreen aeromedical edition professional manual.* Odessa, FL: Psychological Assessment Resources, Inc.

Kehoe, J.F., & Olson, A. (2005). Cut scores and employment discrimination litigation. In F.J. Landy (Ed.), *Employment discrimination litigation: Behavioral, quantitative, and legal perspectives* (pp. 410-448). San Francisco, CA: Jossey-Bass.

King, R.E., Schroeder, D.J., Manning, C.A., Retzlaff, P.D., &Williams, C.A. (2008). Screening air traffic control specialist (ATCS) applicants for psychopathology. *International Journal of Applied Aviation Studies*, *8*, 58-71.

Kokorian, A. (2010). *PILAPT Science fact sheet.* Retrieved from (http://www.pilapt.com/downloads/pilapt_science_fact_sheet.pdf)

Kokorian, A., & Pereira da Costa, M. (2008, September). *On the transportability of a computerized test battery for the selection of pilots.* Paper presented at the International Military Testing Association, Amsterdam, the Netherlands. Retrieved from (http://www.imta.info/PastConferences/Presentations.aspx)

Kokorian, A., Valsler, C., Tobar, J.C., & Riberio, R.B. (2004, October). *Generalisability of the criterion validity for a pilot selection battery.* Paper presented at the International Military Testing Association, Brussels, Belgium.

Kuncel, N.R., Klieger, D.M., Connelly, B.S., & Ones, D.S. (2013). Mechanical versus clinical data combination in selection and admissions decisions: A meta-analysis. *Journal of Applied Psychology*, *98*, 1060-1072.

Lake, K.M. (2011, April). *A critical analysis of the Airline Safety and Federal Aviation Administration Extension Act of 2010.* (Paper No. 84). Retrieved from Southern Illinois University Research Papers (http://opensiuc.lib.siu.edu/gs_rp/84)

Malos, S. (2005). The importance of valid selection and performance appraisal: Do management practices figure in case law? In F.J. Landy (Ed.), *Employment discrimination litigation: Behavioral, quantitative, and legal perspectives* (pp. 373-409).

Manning, C.M. (Ed.) (2000). *Measuring air traffic controller performance in a high-fidelity simulation.* (Report No. DOT/FAA/AM-00/2). Washington, DC: Federal Aviation Administration Office of Aviation Medicine. Retrieved from (http://www.faa.gov/data_research/research/med_humanfacs/oamtechreports/2000s/media/00_02.pdf)

Manning, C.A., Della Rocco, P.S., & Bryant, K.D. (1989). *Prediction of success in FAA air traffic control field training as a function of selection and screening test performance.* (Report No. DOT/FAA/AM-89/6). Washington, DC: Federal Aviation Administration Office of Aviation Medicine. Retrieved from

(http://www.faa.gov/data_research/research/med_humanfacs/oamtechreports/1980s/media/AM89-07.pdf)

Manning, C.M., & Heil, M.C. (2001). The relationship of FAA archival data to AT-SAT predictor and criterion measures. In R.A. Ramos, M.C. Heil, M.C., & C.A. Manning. (Eds.), *Documentation of validity for the AT-SAT computerized test battery: Vol. II* (pp. 49-61). (Report No. DOT/FAA/AM-01/6). Washington, DC: Federal Aviation Administration Office of Aviation Medicine. Retrieved from (http://www.faa.gov/data_research/research/med_humanfacs/oamtechreports/2000s/media/0106.pdf)

Martinussen, M. (1996). Psychological measures as predictors of pilot performance: A meta-analysis. *International Journal of Aviation Psychology*, *6*, 1-20.

Martinussen, M. (2014). Commentary on the Article by Hoermann and Goerke: Do pilots need social competence? *International Journal of Aviation Psychology*, *24*, 32-35.

Martinussen, M. (2016). Pilot selection: An overview of aptitude and ability assessment. In R. Bor, C. Eriksen, M. Oakes, & P. Scragg (Eds.), *Pilot mental health assessment and support: A practitioner's guide* (pp. 51-67). New York, NY: Routledge.

Maschke, P. (1987). *Temperment Structure Scales.* (Tech. Rep. No. ESA-TT.1069). Oberpfaffenhofen, Germany: European Space Agency.

Maschke, P. (2004). Personality evaluation of applicants in aviation. In K.-M. Goeters (Ed.), *Aviation psychology: Research and practice* (pp. 141-151). Burlington, VT: Ashgate.

Maurino, D.E., & Murray, P.S. (2010). Crew resource management. In J.A. Wise, V.D. Hopkin, & D.J. Garland (Eds.) *Handbook of aviation human factors (2d Ed.)* (pp. 10-20). Boca Raton, FL: CRC Press.

McGee, M. (2015). *Air transport pilot supply and demand: Current state and effects of recent legislation.* Unpublished Ph.D. dissertation, Pardee RAND Graduate School, Santa Monica, CA. Retrieved from (http://www.rand.org/content/dam/rand/pubs/rgs_dissertations/RGSD300/RGSD351/RAND_RGSD351.pdf)

McPhail, S.M. (2007). *Alternative validation strategies: Developing new and leveraging existing validity evidence.* San Francisco, CA: Jossey-Bass.

McPhail, S.M., Jeanneret, P.R., McCormick, E.J., & Mecham, R.C. (1998). *Position Analysis Questionnaire: Job Analysis Manual* (3rd Ed.). North Logan, UT: PAQ Services, Inc.

Mesorosova, K. (2013, November). *Design of an ab-initio pilot selection process with reduced cultural and language bias.* Paper presented at the International Conference on Air Transport (INAR 2013), Bratislava, Slovakia. DOI: 10.13140/2.1.4661.7283 (https://www.researchgate.net/publication/268802402_Design_of_an_Ab-initio_Pilot_Selection_Process_with_Reduced_Cultural_and_Language_Bias)

Moak, L. (September 18, 2014). Opinion: The pilot shortage myth. *Aviation Week & Space Technology*. Retrieved from (http://aviationweek.com/commercial-aviation/opinion-pilot-shortage-myth)

Nickels, B., Bobko, P., Blair, M., Sands, W., & Tartak, E. (1995, January). *Separation and Control Hiring Assessment (SACHA) final job analysis report.* (Contract Deliverable CDRL-007A under FAA contract DTFA01-91-C-00032). Washington, DC: Federal Aviation Administration Office of Personnel.

O'Connor, P., Flin, R., Fletcher, G., & Hemsley, P. (2002). Methods used to evaluate the effectiveness of flightcrew CRM training in the UK aviation industry. *Human Factors and Aerospace Safety*, *2*, 235-255.

Older, H.J. & Cameron, B.J. (1972, February). *Human factors aspects of air traffic control.* (NASA CR-1957). Washington, DC: National Aeronautics and Space Administration. Retrieved from (http://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/19720011452_1972011452.pdf)

People Technologies, Inc. (2015). *PILAPT®*. Retrieved from (http://www.pilapt.com/default.htm)

Pierce, L.G., Broach, D., Byrne, C.G., & Bleckley, M.K. (2014). *Using biodata to select air traffic controllers.* (Report No. DOT/FAA/AM-14/8). Washington, DC: Federal Aviation Administration Office of Aerospace Medicine. Retrieved from (http://www.faa.gov/data_research/research/med_humanfacs/oamtechreports/2010s/media/201408.pdf)

Prince, C. & Salas, E. (1999). Team processes and their training in aviation. In D.J. Garland, J.A. Wise, & V.D. Hopkin (Eds.), *The Handbook of aviation human factors* (pp. 193-214). Mahwah, NJ: Lawrence Erlbaum Associates.

Pulakos, E.D. (2005). *Selection assessment methods: A guide to implementing formal assessments to build a high-quality workforce.* Alexandria, VA: SHRM Foundation. Retrieved from (https://www.shrm.org/hr-today/news/hr-magazine/Documents/assessment_methods.pdf)

Pyburn, K.M., Ployhart, R.E., & Kravitz, D.A. (2008). The diversity-validity dilemma: Overview and legal context. *Personnel Psychology*, *61*, 143-151.

Ramos, R.A., Heil, M.C., & Manning, C.A. (Eds.). (2001a). *Documentation of validity for the AT-SAT computerized test battery: Vol. I.* (Report No. DOT/FAA/AM-01/5). Washington, DC: Federal Aviation Administration Office of Aviation Medicine. Retrieved from (http://www.faa.gov/data_research/research/med_humanfacs/oamtechreports/2000s/media/0105.pdf)

Ramos, R., A., Heil, M.C., & Manning, C.A. (Eds.). (2001b). *Documentation of validity for the AT-SAT computerized test battery: Vol. II.* (Report No. DOT/FAA/AM-01/6). Washington, DC: Federal Aviation Administration Office of Aviation Medicine. Retrieved from (http://www.faa.gov/data_research/research/med_humanfacs/oamtechreports/2000s/media/0106.pdf)

Regional Airline Association. (2014, February 28). *RAA urges Congress and FAA to immediately address pilot supply challenge.* Retrieved from (http://www.raa.org/news/news.asp?id=187064&hhSearchTerms=%221500+and+hour+and+rule%22)

Resource Group. (2015a). *ADAPT full system.* Retrieved from (http://www.resourcegroup.co.uk/what-we-do/symbiotic-performance-solutions/services/human-performance-assessment-and-screening/full-adapt/)

Resource Group. (2015b). *Future Adaptive Screening Tool (FAST).* Retrieved from (http://www.resourcegroup.co.uk/what-we-do/symbiotic-performance-solutions/services/human-performance-assessment-and-screening/fast/)

Roscoe, S., Corl, L., & LaRoche, J. (1997). *Predicting human performance.* Pierrefonds, Quebec, Canada: Helio Press.

Roscoe, S., Corl, L., & LaRoche, J. (1999). *Keeping the picture: The measurement of flow control.* Pierrefonds, Quebec, Canada: Helio Press.

Ryan, A.M., & Ployhart, R.E. (2000). Applicants' perceptions of selection procedures and decisions: A critical review and agenda for the future. *Journal of Management*, *26*, 565-606.

Sackett, P.R., De Corte, W.D., & Lievens, F. (2008). Pareto-optimal predictor composite formation: A complementary approach to alleviating the selection quality/adverse impact dilemma. *International Journal of Selection and Assessment*, *16*, 206-209.

Salas, E., Fowlkes, J.E., Stout, R.J., Milanovich, D.M., & Prince, C. (1999). Does CRM training improve teamwork skills in the cockpit?: Two evaluation studies. *Human Factors*, *41*, 326 – 343.

Salas, E., Wilson, K. A., Burke, C. S., & Wightman, D. C. (2006). Does crew resource management training work? An update, an extension, and some critical needs. *Human Factors*, *48*(2), 392-412.

Scheck, M.F. (2006). *Multi crew pilot license: How many hours of flight training does it take to become an airline pilot?* Unpublished master's thesis, Embry-Riddle Aeronautical University, Dayton Beach, FL. Retrieved from (http://www.vcockpit.de/fileadmin/dokumente/themen/arbeitsgruppen/releasable_version_grp_12Oct.pdf)

Schnurman, M. (2015, October 24). From Dallas' perch, consolidation looks sweet. *Dallas Morning News*, http://www.dallasnews.com/business/columnists/mitchell-schnurman/20151023-airline-mergers-paying-off.ece

Schroeder, D. J., Broach, D., & Young, W.C. (1993). *Contributions of personality measures to predicting success of trainees in the Air Traffic Control Nonradar Screen program.* (Report No. DOT/FAA/AM-93/4). Washington, DC: Federal Aviation Administration Office of Aviation Medicine. Retrieved from (http://www.faa.gov/data_research/research/med_humanfacs/oamtechreports/1990s/media/AM93-04.pdf)

Schroeder, D. J., & Dollar, C.S. (1997). *Personality characteristics of Pre/Post Strike air traffic control applicants.* (Report No. DOT/FAA/AM-97/17). Washington, DC: Federal Aviation Administration Office of Aviation Medicine. Retrieved from (http://www.faa.gov/data_research/research/med_humanfacs/oamtechreports/1990s/media/AM97-17.pdf)

Sells, S.B., Dailey, J.T., & Pickrel, E.W. (Eds.). (1984). *Selection of air traffic controllers.* (Report No. DOT/FAA/AM-84/2). Washington, DC: Federal Aviation Administration Office of Aviation Medicine. Retrieved from (http://www.faa.gov/data_research/research/med_humanfacs/oamtechreports/1980s/media/AM84-02.pdf)

Smith, R.C. (1971). *Personality assessment in aviation: An analysis of the item ambiguity characteristics of the 16PF and MMPI.* (FAA Report No. DOT/FAA/AM-71/35). Washington, DC: Federal Aviation Administration Office of Aviation Medicine. Available from (https://www.faa.gov/data_research/research/med_humanfacs/oamtechreports/1970s/1971/)

Smith, G., Bjerke, E. NewMyer, D., Niemczyk, M., & Hamilton, R.A. (2010). Pilot source study: An analysis of pilot backgrounds and subsequent success in U.S. regional airline training. *International Journal of Applied Aviation Studies*, *10*, 73-96.

Smith, G.M., Herchko, D., Bjerke, E., Niemczyk, M., Nullmeyer, R., Paasch, J., & NewMyer, D.A. (2013). The 2012 pilot source study (Phase III): Response to the pilot certification and qualification requirements for air carrier operations. *Journal of Aviation Technology and Engineering*, *2*(2), 13-23.

Society for Industrial and Organizational Psychology (SIOP). (2003). *Principles for the validation and use of employee selection procedures* (4th ed.). Bowling Green, OH: Author.

Sommer, M., Olbrich, A., & Arendasy, M. (2004). Improvements in personnel selection with neural networks: A pilot study in the field of aviation psychology. *International Journal of Aviation Psychology*, *14*, 103-115.

Sommer, M., Häusler, J., Koning, A.J., & Arendasy, M. (2006, October). *Validation of the Dutch Airforce test battery using artificial neural networks.* Paper presented at the 48th International Military Testing Association, Kingston, Canada. Retrieved from (http://www.imta.info/PastConferences/Presentations.aspx)

Stahlberg, G., & Hörmann, H-J. (1993). *International application of the DLR test system: Validation of the pilot selection for IBERIA.* (DLR-Forschungsbericht. 93-42, 74 S). Hamburg, Germany: Deutsches Zentrum für Luft- und Raumfahrt (DLR).

Stark, D.W. (1994). *The validity of cognitive and noncognitive predictors over time.* Unpublished doctoral dissertation, the University of Tulsa, Oklahoma.

Stead, G. (1991a). A validation study of the QANTAS pilot selection process. In E. Farmer (Ed.), *Human resource management in aviation* (Vol. 1, pp. 3-18). Burlington, VT: Ashgate.

Stead, G. (1991b). Qantas pilot selection procedures: Past to present. In Farmer, E. (Ed.), *Human resource management in aviation* (Vol. 2, pp. 176-181). Brookfield, VT: Avebury.

Syversen, J.L. (2012, January). *The SIAP Norway pilot selection model and "Guidance Material and Best Practices for Pilot Aptitude Testing" (IATA, June 2010).* Nesoddtangen, Norway: SIAP Norway AS.

The Age 60 rule: Hearing before the Senate Committee on Commerce, Science, and Transportation, 107th Congress (2001) (testimony of L.Nicholas Lacey). Retrieved from (http://testimony.ost.dot.gov/test/pasttest/01test/Lacey1.htm)

Thompson, D.E., & Thompson, T.A. (1982). Court standards for job analysis in test validation. *Personnel Psychology*, *35*, 865-874.

Tonowski, R. (2015). Test validation pitfalls. In Hanvey, C., & Sady, K. (Eds.). (2015). *Practitioner's guide to legal issues in organizations* (pp. 49-83). New York, NY: Springer.

Trites, D.K. (1961). Problems in air traffic management: I. Longitudinal prediction of effectiveness of air traffic controllers. *Aerospace Medicine*, *32*, 1112-1118.

Trites, D.K. (1964). Problems in air traffic management: VI. Interaction of training-entry age with intellectual and personality characteristics of air traffic control specialists. *Aerospace Medicine*, *35*, 1184-1194.

Trites, D.K. & Cobb, B.B. (1962, February). *Problems in air traffic management: III. Implications for training and job performance of air traffic controllers.* (Report No. DOT/FAA/AM-62/3). Washington, DC: Federal Aviation Administration Office of Aviation Medicine. Retrieved from (http://www.faa.gov/data_research/research/med_humanfacs/oamtechreports/1960s/media/AM62-03.pdf)

Trites, D.K. & Cobb, B.B. (1964a). Problems in air traffic management: III. Implications of training-entry age for training and job performance of air traffic control specialists. *Aerospace Medicine*, *35*, 336-340.

Trites, D.K. & Cobb, B.B. (1964b). Problems in air traffic management: IV. Comparison of pre-employment, job-related experience with aptitude tests as predictors of training and job performance of air traffic control specialists. *Aerospace Medicine*, *35*, 428-436.

Truxillo, D.M., Bodner, T.E., Bertolino, M., Bauer, T.N., & Yonce, C.A. (2009). Effects of explanation on applicant reactions: A meta-analytic review. *International Journal of Selection and Assessment*, *17*(4), 346-361.

U.S. Bureau of Labor Statistics. (2015). Occupational employment and wages, May 2014 53-2011 Airline pilots, copilots, and flight engineers. *Bureau of Labor Statistics*, Retrieved from (http://www.bls.gov/oes/current/oes532011.htm)

U.S. Government Accountability Office (GAO). (2014, February). *Aviation workforce: Current and future availability of airline pilots.* Retrieved from (http://www.gao.gov/products/GAO-14-232)

VanDeventer, A.D. (June 26, 1981). *Comments on draft report titled "Selection of applicants for the air traffic control occupation."* Oklahoma City, OK: FAA Civil Aeromedical Institute Aviation Psychology Laboratory (AAC-118).

Waugh, G. (2001). Predictor-criterion analyses. In R. Ramos, M.C. Heil, & C.A. Manning (Eds.), *Documentation of validity for the AT-SAT computerized test battery: Volume II* (p. 37-42). (FAA Report No. DOT/FAA/AM-01/6). Washington, DC: Federal Aviation Administration Office of Aviation Medicine. Available from (https://www.faa.gov/data_research/research/med_humanfacs/oamtechreports/2000s/2001/)

Weltin, M., Broach, D., Goldbach, K., & O'Donnell, R. (1992, April). *Concurrent criterion-related validation of air traffic control specialist Pre-Training Screen.* (Final report submitted under FAA contract DTFA01-89-Y-01019). Washington, DC: Federal Aviation Administration Office of the Deputy Administrator.

Westrum, R. & Adamski, A.J. (2010). Organizational factors associated with safety and mission success in aviation environments. In J.A. Wise, V.D. Hopkin, & D.J. Garland (Eds., 2nd Ed.), *Handbook of aviation human factors* (pp. 5-1 – 5-37). Boca Raton, FL: CRC Press.

Wikander, R., & Dahlstrom, N. (2014). *The Multi Crew Pilot License – Revolution, evolution or not even a solution? A review and analysis of the emergence, current situation and future of the Multi-Crew Pilot Licence (MPL).* Lund, Sweden: Lund University School of Aviation. Retrieved from (http://www.tfhs.lu.se/fileadmin/tfhs/doc/The_Multi-Crew_Pilot_License.pdf)

Wikander, R., & Dahlstrom, N. (2016). *The Multi Crew Pilot License Part II: The MPL data – Capturing the experience.* Lund, Sweden: Lund University School of Aviation. Retrieved from

(https://www.tfhs.lu.se\fileadmin\tfhs\doc\The_MPL_Part_II_The_MPL_Data_-_Capturing_the_Experience.pdf)

Wingestad, T. (2005, February). *Selection of offshore rotary wing pilots: Do psychological tests predict simulator performance?* Retrieved from (https://www.wingestad.no/userfiles/sites/files/wingestad_in_english.pdf)

Young, W.C., Schroeder, D.J., Agen, R.A., & Broach, D. (1998, March). *Annotated bibliography: Pilot selection.* (Report prepared for the FAA Aviation Rulemaking Advisory Committee on Pilot Selection). Oklahoma City, OK: FAA Aerospace Human Factors Research Division.

Zierke, O. (2014). Predictive validity of knowledge tests for pilot training outcome. *Aviation Psychology and Applied Human Factors*, *4*(2), 98–105. DOI: 10.1027/2192-0923/a000061

FOOTNOTES

[1]The qualifications for the FCC Restricted Radiotelephone Operator's License (http://wireless.fcc.gov/commoperators/index.htm?job=rr) are

- be a legal resident of (or otherwise eligible for employment in) the United States; and
- be able to speak and hear; and
- be able to keep at least a rough written log; and
- be familiar with provisions of applicable treaties, laws, and rules which govern the radio station you will operate.

APPENDIX A: BEST PRACTICES CHECKLIST FOR EVALUATION OF PILOT SELECTION/SELECTION PROCEDURES

Job Analysis

Was a job analysis conducted and reported?

Does the analysis conform to accepted professional and legal standards?

Does the analysis report conform to the reporting requirements of the *Uniform Guidelines on Employee Selection Procedures* (29 CFR § 1607)?

Predictors

Are the constructs to be assessed by proposed predictors clearly specified (operationally defined)?

Are the constructs to be assessed by proposed predictors based on the job analysis?

What is the reliability of each proposed predictors?

- What is the test-retest reliability over what time interval?
- What is the parallel form reliability based on what index?
- What is the internal consistency based on what index?

What is the construct validity of each proposed predictor?

What evidence is there for convergent/discriminant validity of each proposed predictor?

Criteria

What job performance measures are specified as criteria?

How are the job performance measures (criteria) operationally defined?

What are the performance dimensions are assessed?

When and how are the performance data collected?

What is the reliability of each job performance measure?

For measures based on human observations of performance, what is the degree of inter-rater agreement and inter-rater reliability for the measures?

Validation

Was a formal, company-specific validation study conducted?

What design was used in that validation study, with how large a sample?

How were the predictors sequenced in terms of stages and composites?

What level of statistical power to detect a relationship of what size was afforded by the study design?

What protections against Type I errors (false positives) were used?

What was the raw, uncorrected validity coefficient for each predictor at each stage?

What was the raw, uncorrected validity coefficient for any composite scores at each stage?

What corrections were made for direct restriction of range on each predictor and/or composite?

What corrections were made for indirect restriction of range on each predictor and/or composite?

What corrections were made for criterion unreliability?

What was the corrected validity coefficient for each predictor and/or composite at each stage?

By what statistical procedure was the validity coefficient estimated (simple correlation, multiple linear regression, logistic regression, etc.)?

Decision Rules

How were cut-scores on each predictor and/or composite at each stage determined?

What decision rules were established for making selections (e.g., decisions to allow an applicant to proceed to the next stage and to make a formal bona fide job offer)?

Fairness

What is the adverse impact ratio associated with each decision point?

On what sample sizes were the adverse impact ratios computed?

What alternatives of equal validity but with lesser adverse impact are available, at what cost?

Utility

What changes in training success rates and/or on-the-job performance were estimated on the basis of the validation study (or studies)?

What is the monetary value of those changes in training success rates and/or job performance to the company?

Documentation

Is the development and validation of the pilot selection process documented in accordance with 29 C.F.R. § 1607.15?

Are applicant records maintained for statistical purposes with personally identifying information to enable matching to employment data?

Are applicant flow analyses conducted in accordance with 29 C.F.R. § 1607 and OFCCP guidance?

Are post-hire simulator, check-ride, and/or other performance data collected and matched to applicant records?

On-going Evaluation

What periodic assessments are made of the fairness, validity, and utility of the selection process (from initial application through actual entry-on-duty)? How often?

Other Factors

By what organization were any formal validity studies conducted?

To what degree were the organizations conducting such validity studies independent of the company (air carrier or training school)?

What are the professional credentials of the individuals conducting such validity studies?

APPENDIX B: SAMPLE PILOT TRAINING PROGRAM ENTRY REQUIREMENTS

Program/Air Carrier	**Age**	**Language**	**Education**
CAE / Air Asia	18+	English proficiency	
CAE/Gondia	17-30	Fluent English	Minimum score of 55% in Physics, Mathematics, & English in Indian "10+2" exam
CAE/JetStar Asia	18	Satisfactory English proficiency	Satisfactory completion of Secondary School, and Completed Victorian Certificate of Educatio[1]n (VCE) or equivalent, and VCE Units 3 & 4 - a study score of at least 20 in English (any) and Further Mathematics or equivalent
CAE-OAA / EasyJet	18+	Fluent English	Complete second to 16 grades, UK General Certificate of Secondary Education[2] (GCSE) in 3 Advanced-Levels at grades C or higher, or 5 "Ordinary" level subjects with C or higher
CAE-OAA/ Dragon Air		"Good Command" of English	Completed secondary3 education with "Good Passes" (English, Physics, or Math), or college degree in any discipline (science preferred)
CAE / Tiger Air	18-30	Oral and written English proficiency	Minimum of General Certificate of Education (GCE) at the "Advanced" level or equivalent (may consider 5 GCA 'Ordinary' level courses)
CTC / Monarch	18+	Fluent English	Minimum 5 GCSEs or equivalent and Successfully completed secondary or high school education
CTC / Easyjet	18+		5 GCSEs with minimum of 2 Advanced-level certificates with grades BC or higher, or Bachelor's degree with first-class ("2:1") or second-class ("2.2) honors, or Secondary level education to minimum age 16 with 3 Advanced-levels at grades BBC or above, or Higher (graduate-level) degree such as MSC, MA, MPhil, PhD, or MBA

Program/Air Carrier	Age	Language	Education
CTC / Qatar Airways	18-28	Fluent oral and written English	Minimum of 5 GCSEs or equivalent with C or above, and Completed secondary or high school education
CTC/Virgin Atlantic	18-30	Fluent oral and written English	Minimum of 5 GCSEs or equivalent with C or above, and Completed secondary or high school education
Czech Aviation Academy	17+	Demonstrate Level 4 English proficiency before training begins	High school diploma
CAPA/Sterling	17+		Minimal level pass in English, Physics, and Math
FTE/FLYBE	17+	Fluent English	Completed secondary or high school education, or higher
STAA/Tiger Airways	18+	"Good" proficiency	At least 1 "Advanced" level GCE, or 5 "Ordinary"' level GCEs (in English, Science, and Math)
Alpha Aviation / Sharjah / Air Arabia	17-33		High school diploma or equivalent
Flight Path / Ethiopian Airlines	21-25	Read, write, speak and understand English	BSC in engineering, Computers, or Sciences, and CGPA of 2.75 (3.0 outside Ethiopia) or higher
FTEJerez	17½+	Fluent English	GCSE min. C
FTEJEREZ – British Airways	18-55	Fluent English	5 GCSEs with C or above, or 3 Advanced levels at grades BBC or above, or Bachelor's degree with second (2.2) or first-level (2.1) honors, or Higher Degree (MsC, PhD, or MBA)

Program/Air Carrier	Age	Language	Education
FTEJEREZ – Aer Lingus	18+	Fluent English	4 GCSE with C or above Desired 3rd level qualification or studying
KLM Flight Academy	18+		High school diploma or equivalent
Condor	19-35	German/English	Technical College or University (unspecified if degree required)
Indigo	17-30	Fluent English	Minimum score of 55% in Physics, Mathematics, & English in Indian "10+2" exam

Notes: [1]Victoria Certificate of Education (VCE) is the credential awarded to secondary school students who successfully complete high school level studies (year 11 and 12 or equivalent) in the Australian state of Victoria

[2]General Certificate of Secondary Education (GCSE) is an academic qualification awarded in a specified subject, generally taken in a number of subjects by pupils in secondary education in England and Wales over two years. GCSE can be awarded for "Ordinary" and "Advanced" level subjects. Used primarily in the United Kingdom and other British Commonwealth countries.

[3]Secondary education generally refers to the equivalent of U.S. high school (grades 9 through 12), with primary education referring to grades 1-8.