The Pilot Proficiency Audit of Knowledge, Skills, and Abilities: Adding to the Air Carrier Safety Toolbox

James M. Walters

Embry-Riddle Aeronautical University - Daytona Beach

Follow this and additional works at: http://commons.erau.edu/db-theses

Part of the Management and Operations Commons

Scholarly Commons Citation
THE PILOT PROFICIENCY AUDIT
OF KNOWLEDGE, SKILLS, AND ABILITIES:
ADDING TO THE AIR CARRIER SAFETY TOOLBOX

by

James M. Walters

A Thesis Submitted to the
Extended Campus
in Partial Fulfillment of the Requirements for the Degree of
Master of Aeronautical Science

Embry-Riddle Aeronautical University
Daytona Beach, Florida
November 2002
INFORMATION TO USERS

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleed-through, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.
THE PILOT PROFICIENCY AUDIT
OF KNOWLEDGE, SKILLS, AND ABILITIES:
ADDING TO THE AIR CARRIER SAFETY TOOLBOX

by

James M. Walters

This Thesis was prepared under the direction of the candidate’s Thesis Committee
Chair, Dr. Thomas R. Weitzel, Associate Professor, and the Thesis
Committee Members, Mr. Donald B. Hunt, Associate Professor, and Dr. Marvin L.
Smith, Professor. Upon approval by the Daytona Beach Campus Thesis Committee,
the Thesis was submitted to the Extended Campus and accepted in partial fulfillment
of the requirements for the degree of Master of Aeronautical Science.

THESIS COMMITTEE:

Dr. Thomas R. Weitzel
Committee Chairman

Mr. Donald B. Hunt
Committee Member

Dr. Marvin L. Smith
Committee Member
ACKNOWLEDGMENTS

This thesis was ultimately made possible only by the guidance and encouragement of my thesis committee. I particularly wish to thank Dr. Tom Weitzel for his continual support of this study, and for his insistence upon the pursuit of a quality product.

I am indebted to Captain Hugh Schoelzel for providing a working environment rich in opportunity, inquisitiveness, and laughter. His leadership and understanding have allowed me the time and given me the tools to complete this project. Captains Mike McFarland and Richard Fariello dedicated several months of their time to this study, and the outstanding data obtained was a direct result of their efforts and professionalism.

Finally, to my wife Alana and children Jimmy and Jacquie who have always willingly sacrificed our time together - thank you all!
The correlation between air carrier pilot performance and age, total flight time, total airline flight time, and time in current position has not been fully documented. This study observed 62 F/Os in line operations and graded 74 knowledge, skills, and abilities performance variables, utilizing a five-point Likert scale. Knowledge items scored slightly below the referent, with no improvement over time; skills improved with both flight time and years of service; and abilities declined markedly across all independent variables. Changes to pilot training syllabi and techniques, as well as hiring practices, may be indicated. Integrating Pilot Proficiency Audit data into existing LOSA, AQP, ASAP, and FOQA programs can provide a more robust air carrier safety program.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>ACKNOWLEDGEMENTS</th>
<th>iii</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABSTRACT</td>
<td>iv</td>
</tr>
<tr>
<td>LIST OF TABLES</td>
<td>vii</td>
</tr>
<tr>
<td>LIST OF FIGURES</td>
<td>viii</td>
</tr>
<tr>
<td>ABBREVIATIONS AND ACRONYMS</td>
<td>ix</td>
</tr>
</tbody>
</table>

## CHAPTER

### I INTRODUCTION

- Background ........................................ 1
  - An Air Carrier's Data-based Safety Programs .................. 3
  - Line Oriented Safety Audit (LOSA) .......................... 4
  - Aviation Safety Action Program (ASAP) ......................... 6
  - Flight Operations Quality Assurance (FOQA) .................. 7
  - Advanced Qualification Program (AQP) ......................... 8

- Comparative Analysis of Programs ......................... 10
- The Pilot Proficiency Audit (PPA) ........................ 13
- Statement of the Problem .................................. 14
- Research Questions ...................................... 14
- Delimitations ........................................... 15

### II REVIEW OF THE LITERATURE AND RESEARCH

- Summary of Relevant Data .................................. 16
- Background Studies ....................................... 18
  - Measurement Characteristics ............................. 18
- Observational and Training Studies .................... 20
  - Inter-rater Reliability ................................ 21
  - Referent Reliability .................................... 22
  - Observation Accuracy .................................... 22
  - Rater Agreement ......................................... 23
  - Observer Training ....................................... 24
- Critical Incident Technique ............................... 26
II REVIEW OF THE LITERATURE AND RESEARCH (continued)

Performance Measurement Instrument ........................................ 27
   Event Set ........................................................................... 29
   Observational Worksheet ..................................................... 29
Knowledge, Skills, and Abilities .................................................. 30
The Effect of Age and Experience on Performance ....................... 32
Pilot Error ............................................................................ 33
Previous Air Carrier Performance Assessments ......................... 34

III METHODS AND PROCEDURES .............................................. 36
   The Pilot Proficiency Audit (PPA) .......................................... 37
   Subjects ........................................................................... 38
   Evaluation Procedure ......................................................... 38
   Observers .......................................................................... 40
   Instrument .......................................................................... 42
   Likert Scale ........................................................................ 43

IV ANALYSIS OF DATA ................................................................ 45
   Demographics ...................................................................... 45
   Rating Form Entries ............................................................. 48
      Knowledge ...................................................................... 54
      Skills ............................................................................ 55
      Abilities .......................................................................... 57
   Probationary Pilot Scores ..................................................... 59
   Fleet-wide Studies ............................................................... 63

V CONCLUSIONS & RECOMMENDATIONS ................................ 69
   The PPA ............................................................................ 69
      Knowledge ...................................................................... 69
      Skills ............................................................................ 71
      Abilities .......................................................................... 71
      Probationary Pilot Scores ................................................ 72
      Fleet-wide Studies ............................................................ 72
      Air Safety Programs ........................................................... 73

REFERENCES ............................................................................ 79

APPENDIX ............................................................................... 84
# LIST OF TABLES

<table>
<thead>
<tr>
<th>Table Number</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Types of Data Generated by an Air Carrier’s Safety Programs</td>
<td>11</td>
</tr>
<tr>
<td>2</td>
<td>Pilot Performance Index Variable List</td>
<td>28</td>
</tr>
<tr>
<td>3</td>
<td>Rating Form Graded Entries: Knowledge, Skills, and Abilities</td>
<td>46</td>
</tr>
<tr>
<td>4</td>
<td>Mean KSA Scores for Duplicated Observations of Two F/Os</td>
<td>47</td>
</tr>
<tr>
<td>5</td>
<td>Descriptive Statistics, Independent Variables</td>
<td>47</td>
</tr>
<tr>
<td>6</td>
<td>Descriptive Statistics, Overall Scores and Knowledge, Skills, and Abilities Scores</td>
<td>49</td>
</tr>
<tr>
<td>7</td>
<td>Descriptive Statistics, Knowledge Items</td>
<td>52</td>
</tr>
<tr>
<td>8</td>
<td>Descriptive Statistics, Skills Items</td>
<td>53</td>
</tr>
<tr>
<td>9</td>
<td>Descriptive Statistics, Abilities Items</td>
<td>54</td>
</tr>
<tr>
<td>10</td>
<td>Descriptive Statistics, Probationary Pilots Mean Scores</td>
<td>62</td>
</tr>
<tr>
<td>11</td>
<td>Mean Knowledge, Skills, and Abilities for Each Fleet Type</td>
<td>64</td>
</tr>
</tbody>
</table>
LIST OF FIGURES

1. All mean scores (KSA)/Years of service ......................................................... 50
2. All mean scores (KSA)/Total airline flight time ............................................. 50
3. Selected knowledge items/ Years of service ................................................... 55
4. Selected knowledge items/Flight time in current position ................................ 55
5. Selected skills items/Years of service ............................................................ 57
6. Selected skills items/Flight time in current position ....................................... 57
7. Selected abilities items/Years of service ......................................................... 58
8. Selected abilities items/Total airline flight time ............................................. 58
9. Probationary pilot mean scores/Age ............................................................... 61
10. Probationary pilot mean scores/Total flight time .......................................... 61
11. Traffic lookout means scores/Fleet type ....................................................... 64
12. Mean KSA and total mean/Fleet type ........................................................... 65
13. The system safety loop and air carrier programs ........................................... 76
ABBREVIATIONS AND ACRONYMS

The following is a list of abbreviations and acronyms that appear in the text of THE PILOT PROFICIENCY AUDIT OF KNOWLEDGE, SKILLS, AND ABILITIES: ADDING TO THE AIR CARRIER SAFETY TOOLBOX.

AQP Advanced Qualification Program
ASAP Aviation Safety Action Program
ASRS Aviation Safety Reporting System
ATC Air Traffic Control
ATIS Automatic Terminal Information Service
ATOS Air Transportation Oversight System
BARS Behaviorally-anchored Rating Scales
CAT II Category II
CFIT Controlled Flight Into Terrain
CFR Code of Federal Regulations
CRM Crew Resource Management
ERT Event Review Team
FAA Federal Aviation Administration
FAR Federal Aviation Regulation
FMS Flight Management System
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Full Form</th>
</tr>
</thead>
<tbody>
<tr>
<td>F/O</td>
<td>First Officer</td>
</tr>
<tr>
<td>FOQA</td>
<td>Flight Operations Quality Assurance</td>
</tr>
<tr>
<td>GPS</td>
<td>Global Positioning System</td>
</tr>
<tr>
<td>ICAO</td>
<td>International Civil Aviation Organization</td>
</tr>
<tr>
<td>IEP</td>
<td>Internal Evaluation Program</td>
</tr>
<tr>
<td>ILS</td>
<td>Instrument Landing System</td>
</tr>
<tr>
<td>IOE</td>
<td>Initial Operating Experience</td>
</tr>
<tr>
<td>ISO</td>
<td>International Organization for Standardization</td>
</tr>
<tr>
<td>KSA</td>
<td>Knowledge, Skills, and Abilities</td>
</tr>
<tr>
<td>LOE</td>
<td>Line Operating Environment, or Line Oriented Evaluation</td>
</tr>
<tr>
<td>LOFT</td>
<td>Line Oriented Flight Training</td>
</tr>
<tr>
<td>LOS</td>
<td>Line Operational Simulation</td>
</tr>
<tr>
<td>LOSA</td>
<td>Line Oriented Safety Audit, or Line Operational Safety Audit</td>
</tr>
<tr>
<td>MEL</td>
<td>Minimum Equipment List</td>
</tr>
<tr>
<td>NASA</td>
<td>National Aeronautics and Space Administration</td>
</tr>
<tr>
<td>NRP</td>
<td>National Route Plan</td>
</tr>
<tr>
<td>NTSB</td>
<td>National Transportation Safety Board</td>
</tr>
<tr>
<td>PADB</td>
<td>Program Audit Database</td>
</tr>
<tr>
<td>PMI</td>
<td>Performance Measurement Instrument</td>
</tr>
<tr>
<td>PPDB</td>
<td>Pilot Performance Database</td>
</tr>
<tr>
<td>PPDR</td>
<td>Pilot Performance Description Report</td>
</tr>
<tr>
<td>PNF</td>
<td>Pilot Not Flying</td>
</tr>
<tr>
<td>PPA</td>
<td>Pilot Proficiency Audit</td>
</tr>
<tr>
<td>Acronym</td>
<td>Definition</td>
</tr>
<tr>
<td>---------</td>
<td>------------------------------------------------</td>
</tr>
<tr>
<td>PPI</td>
<td>Pilot Performance Index</td>
</tr>
<tr>
<td>QAR</td>
<td>Quick Access Recorder</td>
</tr>
<tr>
<td>SA</td>
<td>Situational Awareness</td>
</tr>
<tr>
<td>SATS</td>
<td>Small Aircraft Transportation System</td>
</tr>
<tr>
<td>SID</td>
<td>Standard Instrument Departure</td>
</tr>
<tr>
<td>SME</td>
<td>Subject Matter Expert</td>
</tr>
<tr>
<td>TCAS</td>
<td>Traffic Alert &amp; Collision Avoidance System</td>
</tr>
</tbody>
</table>
INTRODUCTION

Background

The history of commercial aviation in the United States (U. S.), the safest and most efficient air transportation system in the world, is a testament to continuous change. From propellers to jets, from simple radio navigation to Global Positioning Systems (GPS), and from "see and avoid" to the Enhanced Traffic Alert & Collision Avoidance System (TCAS) II, virtually every aspect of the industry has been host to remarkable technological development. These improvements have played a major role in the decline of the air carrier accident rate, from about 30 hull loss accidents per million departures in the early 1960s to about 2 per million in 2000. However, as pointed out by the Federal Aviation Administration’s (FAA’s) Safer Skies Initiative, at the current rate the worldwide aviation community can expect one major hull loss accident every week by 2010 (Boeing, 2002).

Safety initiatives are ongoing in many areas, including controlled flight into terrain (CFIT), runway incursions, loss of aircraft control, weather detection and avoidance, and pilot decision making. These efforts are a real step forward in the industry’s stated goal of reducing the fatal accident rate by 80% in the next 5 years. From a larger perspective, however, aviation safety is only achieved by building many layers of overlapping defenses; only by accurately predicting the circumstances that can lead to
accidents and incidents can we hope to build effective barriers or defenses into the aviation system. Because these barriers must depend on personnel and equipment functioning in a reliable and predictable manner, quality management systems have become increasingly important.

In the early years of aviation safety, the mantra seemed to be ‘fly it, crash it, redesign it, fly it, crash it . . .’ resulting in only modest improvements over time. System safety principles, as commonly practiced by most major air carriers now, require that mechanisms be in place to:

1. Identify and manage safety risks where they are predicted to be the greatest.
2. Build and maintain appropriate layers of defense (barriers to risk).
3. Measure the barrier’s ongoing effectiveness.
4. Evaluate system risks for prevention of accidents and mitigation of consequences (Mein, 2002).

This process approach to aviation safety is reflected in other quality management standards, specifically the International Organization for Standardization’s (ISO’s) 9000:2000. These standards describe the processes used by an organization to demonstrate its ability to consistently and reliably provide a product (in this case, safety) that meets customer and applicable regulatory requirements, and provides a basis for continual improvement and business excellence. Several airlines are actively seeking ISO 9000 certification for their flight safety departments.

The FAA has embraced the use of quality management principles in aviation safety by instituting the Air Transportation Oversight System (ATOS). This new
approach to airline certification and surveillance oversight was implemented in 1998.

Unlike earlier traditional oversight methods:

ATOS incorporates the structured application of new inspection tasks, analytical processes, and data collection techniques to the oversight of individual air carriers. This approach enables Flight Standards inspectors to be more effective in the oversight of air carriers by focusing on the most critical safety aspects of an air carrier's operation. As currently applied, ATOS provides a systematic process for conducting surveillance, identifying and dealing with risks, and providing data and analysis to guide the oversight of each carrier. (FAA, 2002, Overview, ¶ 1)

ATOS was developed from the ISO 9000:2000 guidelines, and the two quality mechanisms share many philosophical and "process" attributes. Most important among those, however, is the fundamental requirement for data-based measurement of the airline's safety product (pilot proficiency, training, flight operations, etc.), enhancements to operational systems, and effectiveness of those improvements. Unlike the more traditional methods of subjective safety analysis, accurate and reliable data is the hallmark of a modern air carrier's overall safety program.

An Air Carrier's Data-based Safety Programs

There are four commonly used processes for gathering operational performance data. The Line Oriented Safety Audit is an observational study, the Aviation Safety Action Program relies on voluntary participant reports, Flight Operations Quality Assurance monitors aircraft performance, and the Advanced Qualification Program maintains pilot proficiency through advanced training concepts. These programs all collect useful data; currently it is up to the air carrier to determine how that information is analyzed, and what other data should be gathered to best enhance operational safety.
The Line Oriented Safety Audit (LOSA) is a relatively new safety initiative that uses expert observers to collect data regarding air carrier flight crew performance on normal, scheduled flights. All observations are conducted under strict, non-jeopardy conditions, assuring complete anonymity to the crew. Typically observers are selected from all organizations involved in the audit process, including the LOSA provider, airline management, and the pilots union. Originally begun as an FAA research project in 1996, LOSA was further developed at the University of Texas at Austin and has become the central focus of the International Civil Aviation Organization’s (ICAO’s) current flight safety and human factors program (Maurino, 2002).

Information collected includes those risks and errors external to the flight deck, crew errors, and crew actions to mitigate and manage both risk and error. Five specific types of error are defined in the LOSA: (a) procedural, (b) communication, (c) proficiency, (d) decision making and (e) intentional non-compliance, or violations of regulations (Helmrich, in press). The newest generation of LOSA has incorporated the latest conceptual models of threat and error management. “This change greatly enhanced the usefulness of LOSA for airlines, expanding it from a crew resource management (CRM) audit to one which places skills into perspective as operational threat and error countermeasures” (Helmrich, 2002, p. 7).

As outlined by James Klinect (2002), a successful LOSA study will always have the following 10 characteristics:

1. Observations are made during normal flight operations.
2. Crews volunteer to participate.
3. Data collection is anonymous, confidential and safety-minded.

4. The study has joint (management & union) sponsorship.

5. The observation instrument has appropriate targets – flight crew performance in normal operations.

6. The study employs trusted and trained observers.

7. The study has a trusted data collection site.

8. Data are scrutinized before data analysis, to find inaccuracies and consistency.

9. Data are used to identify areas that need enhancements.

10. The results are given to the pilots. (p.8)

Many LOSAs have been completed at various airlines, producing quantitative views of both external threats and flight crew errors in normal flight operations.

Summarizing the results at three different airlines:

1. External threats and crew errors were pervasive, but differ in their type and frequencies across airlines.

2. The descent/approach/landing phase of flight contained the most threats, errors and consequential outcomes.

3. Intentional errors were the most frequent type, but had the least consequence.

4. Proficiency and operational decision errors were the most difficult for the crews to manage.

5. Automation and checklist usage produced the most common errors, many of which went undetected.
6. CRM behavior is effective in error management. Positive traits include strong leadership, vigilance, and communication skills, as well as effective contingency planning (Klinect, Wilhelm, & Helmrich, 1999, p.687).

Aviation Safety Action Program (ASAP)

In 1976 the National Aeronautics and Space Administration (NASA) introduced the first voluntary self-disclosure program for the reporting of aviation operational errors. That program, the Aviation Safety Reporting System (ASRS), continues to provide valuable data. However, confidentiality and jurisdictional requirements severely limit its ability to correct identifiable aviation hazards.

Aviation Safety Action Programs (ASAPs) are proactive safety initiatives. They are: (a) airline specific, (b) voluntary, (c) self-reporting, and (d) non-punitive. Any employee covered by the airline’s plan (usually pilots and dispatchers) can submit an event report that is either a possible Federal Aviation Regulation (FAR) violation, or a general or specific flight concern. Strict confidentiality is maintained within the program, and with certain limitations, the event reporters are protected from regulatory or certificate action.

An Event Review Team (ERT) composed of airline, labor union, and FAA personnel collaborate on the details of the event, and must reach unanimous consensus on whatever corrective actions to be taken. That action can be either an administrative FAA letter (warning, correction, or no action taken), a response from the ERT directly, or a return of the report to the flight operations department (no FAR violation potential). Recent analysis shows that the ERTs are very effective in reaching that consensus.
through a hierarchy of shared values, a working buffer to exclude
distractions, and sideband communications that build trust. ASAP appears
to be a highly effective cultural mechanism for identifying novel and
subtle hazards, and designing rapid, mutually acceptable corrective
actions. (Ganter, Dean & Cloer, 2000, p. iii)

These programs are proving to be very successful; there are currently 20 airlines with
active ASAP programs (Longridge, 2002).

*Flight Operations Quality Assurance (FOQA)*

Aircraft flight recorders were first introduced in the 1950s as an aid to the aircraft
accident investigator. Mandated by the Civil Aeronautics Administration (later the FAA)
in 1958, these devices are physically mounted in the tail of commercial aircraft, and are
designed to capture and store virtually all of an airplane’s normal flight parameters.
Initially, these recorders were simple, scribe and foil recording mechanisms which used a
hard, pointed stylus to physically imprint markings onto a rotating drum of metallic foil.
It proved to be an effective design, but rather complex, and subject to mechanical failure
and very limited in storage capacity. The latest designs (digital, solid state recorders) can
easily track and store hundreds of in-flight aircraft parameters, over a much greater
sampling time period.

Most parameters are captured every second, but many, especially engine
indications (temperatures, pressures, etc.), can be sampled at much greater rates if
necessary. This is particularly helpful for maintenance quality assurance. Flight
parameter information is still used in the investigative capacity, but the newest recording
systems have been supplemented with additional hardware, the Quick Access Recorder
(QAR), which provides immediate and constant availability, and thus monitoring, of the aircraft data.

An airline’s Flight Operations Quality Assurance (FOQA) program is responsible for reviewing all this data on a routine basis, gathering valuable information on crew actions and aircraft operations. This provides “a valid, empirical record of crew inputs to controls, aircraft navigation and engine parameters . . . contributing to safety by identifying error inducing environmental conditions at certain airports and trends in crew behaviour over time” (Helmrich, 2002, p. 6). The information can be used to improve effectiveness of operational procedures, maintenance and engineering procedures, other safety programs, and pilot training. The concept is controversial – but as Norman Mineta, U.S. Secretary of Transportation, recently stated, “Aviation safety must be taken to a new level, a level that won’t be achieved by doing things the way they have been done in the past” (Mineta, 2002).

The key to FOQA acceptance with pilots is that the data is collected within strict non-jeopardy guidelines, providing immunity even if evidence of a procedural violation is found. Because the data is protected from public disclosure by FAA order (49 U.S.C. 40123), air carriers voluntarily share this and other safety information with the FAA.

**Advanced Qualification Program (AQP)**

The introduction of the Advanced Qualification Program (AQP) by the FAA in 1991 fundamentally changed the traditional concepts of flight crew training. The new, alternative methods of qualifying and certifying pilots under Code of Federal Regulations (CFR) Parts 121 and 135 are strictly proficiency-based, and are airline specific. An air
carrier’s participation is voluntary, but the benefits realized from implementation of AQP include: training tailored to the carrier’s operation, cost reductions due to more efficient scheduling of training, and enhanced flight crew proficiency (Kaempf & Klinger, 1993).

As flexible as the program may be, however, there are a few mandated FAA requirements:

1. The program must be airplane specific to each make, model, or series of aircraft, and include basic indoctrination, qualification, and continuing qualification curriculums for every duty position.

2. The training should be conducted with a full crew complement (Captain and First Officer), and must include a Line Operational Evaluation (LOE). The LOE consists of a simulator-based, real-time flight scenario designed to address specific flight operations and CRM skills.

3. Instructor/evaluators must be trained in specific strategies to assure reliable pilot performance and standardization verification. They must also collect proficiency data on students, for internal evaluation of curriculum development and to supply to the FAA (Longridge, 1997). Data that is collected is routinely stored in two databases: a Program Audit Database (PADB) and a Pilot Performance Database (PPDB). The PPDB contains all pilot initial qualification, LOE, Initial Operating Experience (IOE), and continuing qualification results (Holt, 1997).

Pilot performance as assessed by the AQP is done in a full flight simulator. The various checks consist of Maneuver Validation and Line Oriented Flight Training (LOFT). Maneuver Validation, or “first look,” consists of a series of specific, predetermined, maneuvers (engine failure at $V_1$, stalls, etc.). The LOFT employs a series of
“event sets” which make up a complete flight scenario, to be flown in real time by a complete flight crew. The emphasis is on CRM, decision making, and teamwork. The importance of CRM in the overall context of AQP is evident in recent efforts to “remove the separation between CRM and technical training [flight skills] to encourage crews to treat utilization of good CRM skills as little different from employing effective technical skills” (Mangold & Neumeister, 1995, p. 556).

The collection of valid, reliable data is critical to the success of this, or any other proficiency-based program. A complete discussion of the principles of accurate AQP data collection can be found in later in this report.

Comparative Analysis of Programs

These four primary data-gathering safety programs provide air carriers operational information essential for monitoring and improving safety. Each has advantages over the other, and each one has certain limitations inherent to the program. Table 1 is a matrix that illustrates the individual air carrier safety programs and the pilot proficiency information that each was designed to measure.

Pilot performance can be classified into three basic areas: Knowledge, skills, and abilities. For purposes of this thesis, the following definitions are assumed:

1. Knowledge – those facts, policies, and procedures that are taught to a student pilot by the airline, and for which the pilot has the responsibility to remember. Examples would be company policies as contained in the Administrative Manual and aircraft operating limitations as contained in the Flight Handbook.
2. Skills — those flying skills that are taught to the student pilot by the airline during initial training, including all techniques applicable to a particular type of aircraft to be flown at the airline. Not included would be normal “stick and rudder” skills that the pilot would have brought with him to the job. Examples would be a crosswind landing or engine start in the aircraft normally flown.

3. Abilities – those attributes that are individual, that are not taught by the airline, and are reflected in all aspects of a pilot’s professional career. Examples include workload management, decision making, attitude and personal appearance.

All three categories can be observed as either crew-based or individual attributes. Both knowledge and skill can be further refined into those events that are rehearsed and those that are observed during normal line operations. Abilities, as individual traits, are always

Table 1

*Types of Data Generated by an Air Carrier’s Safety Programs*

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Personnel</th>
<th>Scenario</th>
<th>FOQA</th>
<th>ASAP</th>
<th>LOSA</th>
<th>AQP</th>
<th>PPA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Knowledge</td>
<td>Crew</td>
<td>Rehearsed Line ops</td>
<td>□</td>
<td>■</td>
<td>□</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Individual</td>
<td>Rehearsed Line ops</td>
<td>□</td>
<td></td>
<td>■</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Skills</td>
<td>Crew</td>
<td>Rehearsed Line ops</td>
<td>□</td>
<td>■</td>
<td></td>
<td>□</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Individual</td>
<td>Rehearsed Line ops</td>
<td>■</td>
<td></td>
<td>■</td>
<td>□</td>
<td>■</td>
</tr>
<tr>
<td>Abilities</td>
<td>Crew</td>
<td>Rehearsed Line ops</td>
<td></td>
<td>■</td>
<td>□</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Individual</td>
<td>Rehearsed Line ops</td>
<td>□</td>
<td></td>
<td>■</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

■ Denotes primary area of observation
d Denotes secondary area of observation
present, and are assumed to be constant in any situation. The focus of any observational program can be to detect negative performance, i.e. errors (FOQA and ASAP), or to detect positive performance (LOSA). Two (AQP and the Pilot Proficiency Audit, or PPA) can do both.

FOQA embodies the simplest concepts of any of the safety programs. Aircraft performance data is collected, and the flight skills of an individual (person flying) or to a lesser extent the flight crew (decision making) can be documented. Because there is no observer “intrusion” into the flight deck, it provides a very reliable, although limited, view of a pilot’s flight skills during normal airline operations. In a sense, the “what” happened is evident, but not the “why” it happened.

ASAP provides some of the same information, but can be distorted due to the reporting mechanism. Information regarding piloting skills and knowledge (individual and crew-based) may be discernable. But since all of the data is self-reported, the validity and reliability of the reporting may be suspect. Its strengths lie in the volume of data generated, trending of improvements initiated, and overall awareness of safety issues; not the monitoring of pilot knowledge, skills, and abilities.

LOSA is an observational process, and provides excellent information regarding flight crew performance in a normal line operational environment. Crew-based knowledge, skills, and abilities can be reliably documented. An individual’s knowledge and skill level can also be observed, but the purpose of the audit is almost exclusively to assess the CRM, decision-making, and threat and error management of a line flight crew.

AQP is exclusively a training program, and is not intended to capture any performance metrics during normal line operations. The flight maneuvers that are
assessed during AQP are performed in a simulator, and are known and practiced for beforehand. The LOFT sessions, while similar to real world conditions, are simulated, and trainees are aware of scenario details prior to the training session. AQP, then, provides an excellent look at the performance of a flight crew (both knowledge and skill) in a rehearsed, simulated environment. Through “first look” maneuvers, it can also detect individual flying skills, and to a lesser extent, individual knowledge.

A complete air-carrier safety program must provide reliable data in all pilot performance categories; knowledge, skills, and abilities. Referring to Table 1, it becomes evident that there are some areas that are not documented utilizing only these four programs. Information regarding individual flying skills in a line environment should be supplemented by observation of positive events. Awareness of an individual pilot’s knowledge in the line environment is almost completely lacking, and there is no method by which an airline can routinely and reliably assess an individual’s abilities.

The industry, however, is aware of this shortfall. As stated in a recent data basing study, “In addition to the required AQP databases, possible carrier databases that would give relevant information include . . . a database focused on pilot Knowledge, Skills, and Abilities (KSA).” Furthermore, “the pilot KSA database could include demographic data such as pilot experience, (total hours, hours in type), individual skill/ability assessments, and other measures such as [pilot’s] attitudes” (Holt, 1997, p. 58).

The Pilot Proficiency Audit (PPA)

To observe and document those categories of pilot performance that are not currently assessed, an airline-wide PPA could be performed. The PPA, in some ways
similar to a LOSA audit, would utilize trained pilot observers to monitor line operations in a confidential, non-jeopardy environment. Emphasis would be on individual (not crew) performance, and would include pilot interviews as well as observations. Consistency of grading would be enhanced by utilizing a standard instrument, a simple Likert scale, advanced training, and a limited number of observers. The recording instrument should be easy to use, clear in its intent, and familiar to the observers.

Certain demographics of the subject population are a matter of record. Date of hire, age, total flight time, total flight time with the air carrier, and time in current aircraft are known for all air carrier pilots. This data could be correlated with specific scores to determine pilot performance trends. The results could be used to improve both ground and flight training techniques and syllabi; to support new training or monitoring efforts; and could be valuable in the new-hire pilot selection process.

Statement of the Problem

The correlation between pilot performance and selected personal and professional factors has not been fully documented. Data generated as a result of a properly designed pilot proficiency study could prove very valuable as yet another component in the overall data-collection processes of an air carrier’s safety program.

Research Questions

The research questions that were addressed were: “Within an air carrier’s pilot population, (a) are their KSAs directly related to the pilot’s age, company longevity, total flight time, or time in current position?” and (b) “Could an individual pilot’s KSA data be
integrated effectively into the air carrier’s overall data collection processes to enhance flight safety?”

Delimitations

This PPA was conducted at an air carrier that operated four different types of large, transport category jet aircraft, with a seniority list of approximately 2,500 pilots. All observations were of two-pilot crews operating in controlled airspace under CFR Part 121 regulations, and focused specifically on the performance of the air carrier’s First Officers (F/Os). The application of the study methodology, conclusions, and/or recommendations to any other aviation operator or air carrier may or may not be appropriate, due to the extensive variations of airline size, aircraft types, operating philosophies, and cultures.
CHAPTER II
REVIEW OF RELEVANT LITERATURE AND RESEARCH

Summary of Relevant Data

One could assume that since on-the-job performance of all pilots, whether civil, airline, or military, is constantly and repeatedly measured, that applicable and timely reference material concerning those measurements would be readily accessible. Unfortunately, that is not the case. Virtually all recent pilot performance studies have focused on the human performance (human factors) issues of workload, situational awareness (SA), and vigilance, all of which are all vitally important characteristics of a successful, professional airline pilot. The Pilot Proficiency Audit (PPA) acknowledges their significance through the measurement of CRM, SA, workload management, decision making, and cockpit discipline.

Of greater importance to this review, however, are the physical manipulation of aircraft controls in a line environment, the knowledge of policies and procedures as taught by the air carrier, and the individual personality traits of each observed pilot. It is in these areas that data is scarce. As stated in a 1997 report, “There is little research in the commercial aviation domain examining the pilot’s overall flight performance in assessment events like the LOE, maneuver validation and line checks” (Boehm-Davis, Holt, & Hansberger, 1997, p. 462). Even in those areas that are usually grouped under the
CRM banner, "The research base for the use of these behaviours is... quite thin"  

There are studies in other areas that have proven valuable for purposes of this research. They can be grouped into the following general categories:

1. Background studies in other areas related to pilot performance, and human performance in general.

2. The design of proven observational and training exercises, including use of appropriate rating scales, rater-reliability, and other methods and techniques.

3. Surveys that have identified those pilot traits, or KSAs, proven necessary for success in the industry.

4. Studies that focus on the effects of age and experience on pilot performance, including analysis of those pilot characteristics commonly identified with "pilot-error" accidents.

5. Pilot assessment studies carried out by the air carrier in the past. This PPA measures the performance of a select group of professional individuals, those employed by one particular air carrier. As such, the only reference material corresponding exactly to that population are previous performance assessments done by that air carrier. Those studies are not available to the public, and as such will not be referenced specifically in this literature review, but certain aspects of those programs have been researched to assure a more complete understanding of performance-rating processes previously undertaken by this air carrier.
Background Studies

The Human Performance Measures Handbook (Gawron, 2000) was originally developed to assist researchers in the selection of appropriate measures in the evaluation of the interaction between humans and machines. Again, the focus is on traditional human factors, but there are several important investigative aspects relevant to a PPA. Those issues outlined in the Handbook that are directly applicable to an experimental study of pilot performance include:

1. Precise definition of the question, description of the independent and dependent variables, and all experimental conditions.
2. Documentation of "qualifiers", or variables that qualify or restrict the generalizability of the results.
3. Matching of subjects with real-world, end users.
4. Selection of proper performance measures, which must be relevant, reliable, valid, quantitative, and comprehensive.
5. Use of a statistically appropriate sample size.
6. Selection of proper data collection and data recording equipment and processes.

Measurement Characteristics.

Of particular importance is step number 4. The pilot performance measures observed and recorded in a PPA should be:
1. Relevant – The KSAs that are the focus of the audit must be the ones that need to be measured, and those that will provide the greatest potential benefit for improvements in safety.

2. Reliable – A reliable measurement is one that is repeatable.

   If one measures the same behavior in exactly the same way under identical circumstances, the same value of the metric should result. In human performance measurement, however, individual differences among human operators, decision-makers, and maintainers occur; even the same person may respond to successive trials differently. (Rehmann, 1995, p. 4)

   In the case of observational studies, the observers are the key to reliability, and for pilot performance audits, inter-rater reliability is essential (this issue is discussed in detail later in this chapter).

3. Valid – Validity refers to measuring what the study really intends to measure, and being appropriate to use for the intended purpose. There are five types of validity usually associated with human performance: face, concurrent, content, construct, and predictive. For purposes of this PPA, the two most important are: (a) face validity - subject matter experts (SMEs) determine that a particular measurement represents the particular pilot performance important to accomplish the task, and (b) predictive validity – the measures being observed in the study should be representative and predictive of the pilot’s performance in the real world. Validity in general is closely tied to reliability (Rehmann, 1995).

4. Quantitative measurements – A quantitative measurement provides an estimate of the size of the difference between experimental conditions, or in the case of a PPA, performance observations. Various rating scales can be
employed, including nominal (identifies differences, with no notion of order, magnitude or size), ordinal (in order, but differences in position not comparable), interval (equal distances between values), or ratio (possesses a true zero) (Morrow, Jackson, Disch, & Mood, 1995).

5. **Comprehensive** – A study that is comprehensive measures all aspects of a pilot’s performance. Observing and recording all possible measurements will be easier and more efficient than repeating the study at a later date to gather data that was initially overlooked. Measurements must also gather sufficient detail to permit a meaningful analysis.

6. **Use of a large enough sample size to assure statistically accurate conclusions.**

7. **Sensitivity** – the degree to which a measure will react to changes in the independent variable. The measure itself can be valid and reliable, but may not show a large enough effect to be easily measured.

8. **Intrusiveness** – almost all measures are intrusive to some degree, but the degree to which they alter the task performance will vary. In the case of a PPA, the observer, by the very nature of his/her presence on the flight deck, will create some distraction for the subject. Obviously, for any observational study, less intrusive methods are preferred (Rehmann, 1995).

**Observational and Training Studies**

Historically, objective pilot performance measurement has involved the analysis of deviations from pre-set flight parameter standards (altitude, heading, etc.). These traditional measures usually take the form of subjective, “expert” ratings made by the
flight instructor in a very controlled environment (McIntyre, 1993). More recent developments in pilot training and checking, however, have created new pilot performance measurement opportunities.

The Advanced Qualification Program relies on the observations of qualified check airman in both the LOE and the LOFT in the simulator. While the F/O PPA focuses on operational factors, not training issues, there are a number of related areas that are important to the overall study of pilot performance.

*Inter-rater Reliability.*

As a proficiency-based program, AQP relies on the collection of empirical data to assess and monitor flight crew performance. As such, the collection and analysis of quality data is fundamental to the program’s success (Holt, Johnson, & Goldsmith, 1997).

Ratings of pilot performance based on observed behavior can only be valid if they are reliable; that is, there is consistency in the measurement process. Within the AQP, this consistency is referred to as “inter-rater reliability” and is primarily focused on rater agreement and observation accuracy. Both can be affected by individual evaluator bias, by the measurement tool utilized, by the types of event sets and scenarios employed, and other factors (Baker & Dismukes, 1997). Birnbach and Longridge (1993) noted that because the variance associated with an unreliable measure can not always be distinguished from error, a reliable measure may not always be valid, but an unreliable measure will never be valid. That relationship can be put another way, in that “a valid test is always reliable, but a reliable test is not always valid” (Gay & Airasian, 2000, p. 170).
Referent Reliability

This method of examining observer’s reliability differs from typical rater reliability in that crew behavior is compared to a pre-determined standard of performance, called a “referent,” rather than the evaluators group norm. This referent is sometimes referred to as the “gold standard” for judgements.

Holt, Johnson, & Goldsmith (1997) describe the benefits of using an external referent:

Using a referent to assess evaluator reliability has three important implications. First, high referent-rater reliability results in high rater-reliability. If evaluators are judging crew behavior in accordance with the same set of standards, then they will necessarily agree with another. Second, the distribution of referent scores defines the appropriate distributional characteristics of evaluator scores. Hence, by training to match the referent judgements, evaluators can be trained to match the mean, variance, and skewness of the referent distribution. Third, a major advantage of referent reliability over inter-rater reliability is that the problem of an incorrect group norm for judgements is avoided. That is, in the case where the majority of raters are rating incorrectly, the other raters will not be trained to the incorrect standard. (p. 917)

Observation Accuracy

Under AQP, a check airman observes what the flight crew says and does during the LOFT or LOE scenario. Because these observations form the basis for the overall
crew performance rating, accuracy is critically important. All raters, therefore, must use the same parameters of "behavioral information" for scoring crews during each event set (Baker & Dismukes, 1997).

While there is substantial reference material on rater agreement, there has been little research regarding observation accuracy in terms of pilot performance. One study (George Mason University, 1996) indicated that raters' average agreement for observed CRM behavior improved only marginally after calibration training, consisting of discussion of inconsistent observations.

Factors that can reduce observational accuracy include the type and number of complex event sets in the scenario, and the fact that the evaluator also has to act as a simulator operator, scenario facilitator, and instructor. The workload associated with performance of these duties can reduce the degree to which accurate observations are made. While the evidence indicates that check airmen acting as evaluators are generally accurate when observing CRM behavior, effective training strategies have yet to be fully developed (Baker & Dismukes, 1997).

**Rater Agreement**

Flight crew performance during an observational PPA is usually documented by the assignment of a grade based on a pre-determined, defined scale. Inter-rater agreement is the extent to which various observers (of the same performance) agree concerning the assigned grade. Several recent studies have shown that raters can achieve reasonable levels of consistency (Brannick, Prince, Salas, & Prince, 1993).
Different rating scales and levels of ratings (overall scenario vs. “event set” ratings) can also affect rater agreement. One recent study concluded that the most reliable results of a PPA could be produced utilizing event set level observations and grading with a 5-point scale (Seamster, Edens & Holt, 1995). For a PPA, then:

Check airmen can demonstrate acceptable levels of both inter-rater agreement and reliability when assessing specific scenario events. In general, check airmen can agree regarding the level of performance demonstrated by the crew with respect to a defined scale (i.e., crew performance was a 3 on a 4-point scale). [However] rater training may need to include more than discussion to ensure that check airmen are accurate and calibrated regarding the carrier’s LOS [Line Operational Simulation] rating scales. (Baker & Dismukes, 1997, p. 454)

Observer Training

Typical rater calibration training in the airline industry consists of a one-day workshop in which the instructors receive required information about the program, discuss various aspects of the air carrier’s rating process, and practice assessing and rating videotaped LOFTs. The evaluators observe a particular flight crew’s taped performance on a given event set, and then individually rate the crew’s performance. After class review, rating discrepancies are discussed among the group to reach consensus. A videotape of the same event, but performed by a different crew, is then rated to determine the group’s level of calibration (Baker & Mulqueen, 1999).

A George Mason University study (Williams, Holt, & Boehm-Davis, 1997) rated the three models of rater training currently used to reduce rater biases and increase reliability and accuracy. First, and best, was “familiarization with behaviorally-based scales prior to rating aids schema formation and accurate categorization of performance” (p. 515). Next, and also important, “frame of reference training (aligning idiosyncratic
standards of a group of raters with organizational standards by establishing a prior frame of reference) has been shown to reduce rater error and increase rater accuracy more effectively than other training approaches” (p. 515). Least effective was the traditional workshop training method.

Effective pilot instructor/evaluator training (identified as industry “best practices”) as described by the American Institutes for Research (Baker & Mulqueen, 1999) should include:

1. A detailed discussion of the LOE or LOFT scenario, both on the overall and event set levels.
2. A detailed discussion of the technical and CRM skills to be assessed.
3. A review of the performance standards for each area to be studied, whether technical or CRM knowledge, skills or abilities.
4. Training to assist raters to be good observers. Discussion should focus on the nature of a good observation, and how to accurately observe a flight crew’s performance.
5. Opportunities to practice and receive feedback on the rating task. This practice should include rating videotaped performances of the specific LOE event sets, over a wide range of crew performance levels (i.e., excellent, good, poor, etc.).
6. Feedback comparing the pilot rater scores with ratings established by “baseline”, or true experts. This feedback is arguably the most import aspect of training, and should include:
(a) The congruency between each observer’s distribution of grades and the group’s distribution.

(b) The degree to which each observer’s mean performance rating differs from the group mean.

(c) The degree to which observers are able to shift their evaluations upward and downward based on improving or decreasing performance.

(d) The degree to which observers can accurately discriminate between crews of varying performance levels (Hamman, Beaubien, & Holt, 1999).

**Critical Incident Technique**

This technique outlines a general set of specifications for observational studies, most of which are appropriate to a PPA. Initially formulated in 1947, the requirements are grouped into three categories: observers, groups being observed, and the behaviors to be observed.

1. **Observers** – Persons doing the observing must be knowledgeable regarding the activity, they must have some relation (personal, professional, etc.) to those being observed, and there must be specific training requirements in order to adequately do the observing.

2. **Groups to be observed** – The study documentation must include a general description of the group, the location, times, and conditions of the planned observations.
3. Behaviors to be observed – Emphasis should be placed on the general type of activity, specific behaviors targeted for observation, and the criteria of the specific behavior that is of relevance and importance to the general aim of the study (Gawron, 2000).

Performance Measurement Instrument

Various performance measurement recording tools have been used over the years, with the Pilot Performance Description Report (PPDR) used by Prophet & Jolley (1969) one of the first. In this instrument, ratings are anchored to specific performance standards, allowing an observer to record desired elements of performance. It also contains a subjective rating scale that allows an evaluator to utilize his/her experience in assessing aviator performance. Numerous studies since that time, including performance investigations by Childs, Spears, and Prophet in 1983, and Kaempf and Blackwell in 1990, have utilized similar tools (Kaempf & Klinger, 1993). All of these studies have demonstrated that proficiency ratings with descriptive scales are more reliable than ratings made without them.

The Pilot Performance Index (PPI) was developed with the aid of subject matter experts (Stein, 1984, p. 20). Essentially, the PPI is a list of performance variables and associated performance criteria that specifically differentiates novice from experienced pilots. Many of the performance variables are directly applicable to the operation of modern transport category, including pitch angle on takeoff, airspeeds, headings, and bank angles. Other variables, such as “course deviation indicator,” “omni bearing
sensor,” and “manifold pressure” are more appropriate for a very generalized pilot population flying less sophisticated aircraft.

Table 2

*Pilot Performance Index Variable List*

<table>
<thead>
<tr>
<th>Takeoff</th>
<th>Climb</th>
<th>Enroute</th>
<th>Descent</th>
<th>Approach</th>
<th>Final Approach</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pitch angle</td>
<td>Heading</td>
<td>Altitude</td>
<td>Heading</td>
<td>Heading</td>
<td>Heading</td>
</tr>
<tr>
<td>Airspeed</td>
<td>Pitch angle</td>
<td>Airspeed</td>
<td>Manifold left</td>
<td>Gear position</td>
<td></td>
</tr>
<tr>
<td>Heading</td>
<td>Heading</td>
<td>Bank angle</td>
<td>Manifold right</td>
<td>Flap position</td>
<td></td>
</tr>
<tr>
<td>Course deviation indicator</td>
<td>Course deviation indicator</td>
<td>Bank angle</td>
<td></td>
<td>Course deviation indicator</td>
<td></td>
</tr>
<tr>
<td>Omni bearing selector</td>
<td>Omni bearing selector</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Kaempf & Klinger’s (1993) Performance Measurement Instrument (PMI) enables observers to document and assess aviator proficiency within the context of specific tasks and conditions. With this tool, pilots are graded in four major flight tasks, loosely grouped as Takeoff, Enroute, Category II (CAT II) Instrument Landing System (ILS) Approach, and Missed Approach, with 79 individual performance elements. In order to be sensitive to various levels of proficiency, but still provide a means for the observer to document unsatisfactory performance, a six point rating scale was developed, with the following verbal anchors: unsatisfactory, acceptable, below average, average, above average, and expert.

Evaluations of the PMI indicate that trained observers can reliably recognize and record specific elements of pilot performance, and can apply the performance rating scales reliably as well. However, a lack of practice with the PMI, observation of both
crewmembers, and difficulty determining the individual pilot contribution to the team performance can produce an undesirable lack of standardization between observers.

**Event Set**

An event set is a group of related flight events which are inserted into a LOFT or LOE scenario for specific training objectives. The set is usually made up of an event trigger (condition that activates the set), distracters (conditions inserted designed to divert the crew’s attention), and supporting events (other events that serve to reinforce the training objectives) (Hamman, Seamster, & Edens, 1995). And while it is understood that the use of appropriate event sets provides the most effective rater training and observational reliability (Seamster, Edens, & Holt, 1995), the LOFT and LOE are training exercises only, and take place in a simulated environment. However, evaluations of line pilot performance could benefit from similar grouping of flight maneuvers into event sets, on both the grading forms and during observer training.

**Observational Worksheet**

As utilized in the AQP, the evaluator’s worksheet is a primary tool for assessing pilot performance. There are a number of features incorporated into the worksheet that serve to improve the overall rating reliability, including:

1. Assessment areas based on event sets, usually subdivided into areas such as flight operations considerations, human performance, situational awareness, and specific maneuvers. “Ratings based on LOE event sets are less confounded by inflation or leniency errors leading to a more reliable

Observation forms used in a recent study (Ikomi, Boehm-Davis, Holt, & Incalcaterra, 1999) were divided into four major areas: Experience and training questions, flight departure phase events, flight cruise phase events, and flight arrival phase events, thus targeting those specific areas of interest in the study.

2. Similarity between the LOE and the air carrier’s line check forms, and in the use of a common rating scale. Familiarity with a common form leads to fewer observer recording errors, and a standard scale used across the full range of air carrier evaluation environments allows the instructors/evaluators to develop pilot performance rating skills based on a consistent set of crew performance standards (Schultz, Seamster, & Edens, 1997).

Knowledge, Skills and Abilities

A common methodology used in personnel research has been the assignment of numerical grades to various categories of human performance. While many grading scales and formats have been used, the selection of appropriate behaviors to observe and grade is critical. It has been noted that it is important that rigorous behavior selections and behaviorally anchored procedures be used in the development of rating scales, regardless of the particular grading format to be used (Landy & Farr, 1980).

Development of appropriate behaviorally-anchored rating scales (BARS) requires the use of just such rigorous procedures. These scales were first proposed in 1954 as an outgrowth of the critical incident technique previously discussed. A recent joint effort of
FlightSafety International and USAirways Express (Bramble, 1997) initially considered 96 job behaviors and 12 broad performance categories as critical in future air carrier pilots. Further research, primarily interviews with pilots and airline management, identified 32 behaviors as “anchors” for nine fundamental behavior scales. Those included: (a) professionalism, (b) awareness, (c) responsibility/initiative, (d) flying skills, (e) communication, (f) judgement/decision making, (g) interactions with passengers, and (h) interactions with flight attendants. While the usefulness of this particular F/O BARS needs further validation, FlightSafety plans to use the scales as a criterion measure in a future pilot selection study.

A later predictive study evaluated general ability, conscientiousness, and stability as predictors of regional airline pilot performance (Bramble, 1998). Job performance was assessed utilizing the nine fundamental BARS scales noted above, resulting in the generation of both “proficiency” and “interpersonal” factors.

A recent study (Jentsch, Bowers, Martin, Barnett, & Prince, 1997) was conducted to identify training needs for those new F/Os about to begin their pilot-not-flying (PNF) duties. Interviews with F/Os, as well as a review of the NASA ASRS, indicated that there is a tendency for F/Os not to adequately monitor and challenge the Captain (one-half of total incidents recorded), while procedural errors accounted for only one-quarter of the errors. The three primary categories of operational errors were: (a) failure to monitor; (b) identification of a problem but choosing not to initiate action; and (c) identification of a problem, recognition of a need to initiate action, but not doing so due to apprehension. During the interview phase of the study, it was documented that a “good” F/O was one who was knowledgeable, demonstrated initiative, was interactive, willing to
communicate, loved flying, had a sense of humor, and presented an overall clean personal appearance.

The Effect of Age and Experience on Performance

For the purposes of a PPA, it would be valuable to know the negative effects (if any) of aging, and the positive effects (again, if any) of flying experience, on both simple flight skills and more complex cognitive tasks. Because of the controversy surrounding the current age-60 retirement rule for U.S. airline pilots, much research has been focused on analyzing the effects of aging on the pilot population, and other studies have focused on experience.

Very recent data has shown that there is a definite correlation between age and specific performance measures. In one study, 100 civilian pilot volunteers age 50 to 69 performed aviation-related tasks in a Frasca flight simulator. Performance indices included staying on course, dialing frequencies, avoiding conflicting traffic, and flying an instrument approach. Nine composite scores were recorded in working memory, visual associative memory, motor coordination, tracking, numerical operations, visual scanning, spatial processing and attribute identification. It was found that pilot age “was significantly correlated with 7 of the 9 factors . . . [with] none of the age relations favoring older pilots” (Taylor, O’Hara, Mumenthaller, & Yesavage, 2000, p. 376).

Specific time-sharing task performance measurements, such as maintaining aircraft stability while navigating, were studied in research conducted several years ago (Tsang, 1995). Of particular interest are those trials that focused on the detrimental effects of aging, and the fact that increased flight experience can mitigate to a limited
degree minor age-related degradation. Experience levels were the sole variable in studies
carried out by the U.S. Air Force (Carretta, Perry, & Ree, 1995). They demonstrated
that flight experience and situational awareness were directly related; the more
experience a pilot had, the better the individual situational awareness was predicted to be.

Pilot Error

Another aspect of pilot performance related to age and experience is decision
making under high workload. Accidents and incidents caused by operational or
behavioral acts of pilots have traditionally been classified as “pilot error.” Review of the
National Transportation Safety Board’s (NTSB’s) records shows that pilot error was a
factor in 38% of all air carrier (CFR Part 121) accidents, 75% of all air taxi accidents
(CFR Part 135), and 85% of general aviation (CFR Part 91) accidents for the period

A comprehensive review of this data was conducted in 2000 (Li, Baker,
Grabowski, Jurek, & Rebok, 2001). That analysis identified certain characteristics of the
pilot-in-command and crash circumstances in those accidents attributed to pilot error.
Importantly, “Age-related variation in the prevalence rates of pilot error was statistically
insignificant for major airline crashes ... [and] total flight time showed an effect on pilot
error only in general aviation crashes” (p. 54). While age and experience play an
important role at the extreme ends of the scales (very young or aged, novice or seasoned
professional), specifically in the general aviation environment, they are generally not
factors in air carrier operations.
Previous Air Carrier Pilot Performance Assessments

The subject air carrier had conducted five flight safety audits between 1976 and 1997. The methodology for each of the audits had been the same; paid volunteers were selected from company management pilots and the line pilot ranks to observe cockpit operations during scheduled line flights. The willing participation and cooperation of the pilots' union had been critical to the success of each audit, and the union's safety representatives contributed substantially to the success of the observational teams. There was no planned method to gather parametric data during any of the five audits; thus no attempt had been made to train (calibrate) the observers, or to assure consistency of the reported findings. Very little, if any, comparative analysis was performed on the results generated from the various audits, or the resultant findings and recommendations.

The aforementioned data gathering method has persisted as a problem in current safety audits of air carrier flight operations. In addition to addressing the need for parametric data, this researcher decided that the pilot performance areas of focus suggested by the earlier audits should include:

1. Knowledge – Flight Operations Manual, logbook, minimum equipment list (MEL), dispatch procedures, weather review, company communications, cockpit discipline and sterile cockpit, and general operating procedures.

2. Skills – Briefings, stabilized approach conformance, preflights, and checklist usage.

3. Abilities – Personal appearance, morale, command ability, attitude, and complacency.
All previous audits had been crew-oriented, rather than individual-pilot-oriented, and all reports had been completely de-identified. No demographic data was available concerning crew experience, company longevity, age, total flight time, or type of aircraft. Despite the absence of previous audits’ parametric data relevant to this study, the flight crews’ familiarity with the auditing processes and the trust instilled by the previous “anonymous” audits tended to enhance the reliability of the data to be gathered during a PPA.
CHAPTER III
METHODS AND PROCEDURES

In early 2001, the safety department of a U.S. air carrier desired to determine the level of operational proficiency among its F/Os. There were a number of factors involved in the decision to obtain accurate pilot proficiency data:

1. The airline had recently been involved in a merger, and the labor integration issues provided ample opportunity for pilot distraction and error.

2. The airline had recently experienced a high turnover rate among its first officers. More than one-half of the 2,000 pilots hired in the previous 7 years had resigned to accept positions with other air carriers, significantly reducing the average experience levels of the remaining pilot workforce.

3. Reports from line check airmen indicated a possible downward trend in first officer operational knowledge, proficiency, and attitude.

4. The airline had only recently completed development of an Internal Evaluation Program (IEP). Exercising the processes and personnel within the IEP could provide beneficial feedback to the embryonic program.

5. There was the potential to compare the data obtained via the airline’s other safety programs (FOQA, AQP, and LOSA, with ASAP under development) with the F/O proposed PPA data for a broader understanding of particular operational safety issues.
6. Finally, there was a traditional aspect of safety that relates to “communication and visibility.” Line pilots needed to know that the airline’s safety professionals are interested and involved in the day-to-day operations, and that there was a reliable mechanism for timely exchange of information. Line qualified auditors interacting with flight crews over a period of at least several weeks would reinforce both of these safety program characteristics.

The air carrier’s executive management approved the development and execution of a pilot proficiency audit, to start on June 1, 2001, and continue for a period of 30 days.

The Pilot Proficiency Audit (PPA)

All air carriers have extensive experience with self-audits, whether in an individual sense (i.e., a check airman rating a particular pilot’s abilities), or in an industry sense, where observers of varied backgrounds scrutinize many different aspects of an airline’s operations (typically a LOSA). A non-traditional F/O PPA could record many of the same parameters, but measuring additional variables of the F/O population would allow the use of correlational techniques not previously considered.

The independent variables of date of hire, age, total flight time, total airline time, and total time in current position were determined to be those that would provide the greatest insight for the study. These variables were available though the host airline’s pilot qualification computer-based tracking system, and were considered independent and reliable.
Subjects

All F/Os of the airline were subject to observation; in the observational flight selection process, no distinction was made regarding fleet (aircraft) type, or operation (domestic vs. international). Personal information was downloaded before the flight was started, but no record was kept of the pilot’s name or company identification number.

Evaluation Procedure

The F/O PPA was conducted in a manner very similar to traditional airline observational studies, i.e. LOSA and LOE. Observers accompanied flight crews on regularly scheduled flights, participating in all phases of the trip, from preflight planning and crew briefings through the final cockpit “end-of-flight” checklist. Most flight crews were familiar with this format from operational safety audits conducted by the airline in previous years. All observations were recorded on an evaluation form, similar in layout and content to the line check airman report form.

All of the air carrier’s flights were subject to auditing. One, two, three and four day flight sequences were included, based on the individual observer’s schedule. Both international and domestic operations were selected, as were flights on all four types of aircraft operated by the carrier. Although the observers started and finished their sequences at only one of the carrier’s “hubs,” F/Os from all of the airline’s crew bases were included in the study.

Observers developed their own schedules, and flights were selected as randomly as possible. However, two other factors were taken into consideration:
1. Both very long and very short flight segments (total flight time) were observed, but obtaining as large a sample size as possible weighted the selection towards shorter flights.

2. If a F/O had already been audited by one observer, then that observer would attempt to not schedule a flight flown by that particular F/O again.

The PPA was not voluntary; a F/O could not refuse the observer access to the cockpit jumpseat. However, all subjects were assured that the study was completely confidential, that no names would be associated with any reports, and that any information used within the airline’s safety program would be further de-identified regarding date and flight number.

Interviews were conducted with the subject F/Os while in cruise flight, when workload permitted. Typically they consisted of discussions regarding the airline’s Operations Specifications, operational policies and procedures, aircraft systems, and performance calculations. The same questions were raised with each F/O. Conversely, the observers encouraged comments and suggestions from the flight crews, all of which were recorded and returned to the Safety Department. The essential elements of these remarks can be found in the Qualitative Data section of Chapter IV.

For a typical correlational study, a sample size of 30 subjects is usually considered accurate (Gay & Airasian, 2000). Although the validity and reliability of the study were believed to be high, a larger sample size would improve the overall results. Working within the constrains of the pilot (observer) contract, it was estimated that each
observer could record two flights on each of the 17 days worked, for a planned total of 68 observations.

Observers

Two observers were selected from a pool of qualified check airmen. One was a current and qualified Captain; the other had recently retired his position as a line Captain. Cumulatively, they had amassed more than 74 years of line operations, and approximately 55,000 hours of flight time, in both narrow and wide body transport category aircraft. Both were qualified as line and simulator instructors, were very familiar with the line check airman grading form and procedures, and had completed extensive AQP initial and recurrent training within the previous 2 years. That training consisted of several video tape sessions, group grading, and group discussion to resolve differences, as described in recent industry reports (Baker & Mulqueen, 1999; Williams, Holt, & Boehm-Davis, 1997).

Prior to initiating the audit, three training sessions were held with the two observers and safety department personnel. Discussions during those meetings centered on the purpose of the study, the methods and the forms to be used, development of techniques that would prove helpful, and verification of inter-rater reliability. Limiting the total number of observers to two assured a high level of agreement regarding performance ratings, particularly given the two individuals’ extensive AQP training and experience.

Several issues that could impact the success of accurate observations were discussed:
1. Previous studies have shown that the presence of another person in an air
carrier cockpit during routine flight operations can be distracting to the point
of disrupting a normal operation. It is known as "intrusiveness," and is
especially true if there is the perception that a check ride is being conducted. It
was extremely important that every effort was made to minimize any altered
flight crew behavior due to the observer’s presence. As a result, techniques
were developed to put the subject F/Os at ease.

2. The behavior of one person in a flight crew can influence the behavior of the
other. Since the focus of the PPA was the F/O, the observers were very
cognizant of the effect certain Captain personalities (domineering, reticent,
etc.) might have on the performance of the F/O, and developed strategies to
minimize the effect on the ratings.

3. It was discussed and understood that personal biases of the observers, whether
personal or professional, must not be allowed to affect the study. If necessary,
an observer would excuse himself from a particular flight; however, all flights
scheduled were completed during the PPA.

The observers were instrumental in developing interview questions, and in
defining the scope of F/O knowledge, skills, and abilities to be graded. These preliminary
meetings were also critical in satisfying the three Critical Incident Technique
specifications for observers, subjects to be observed, and behaviors to be observed.

After two observational flights each, the observer’s preliminary results were
analyzed and discussed. In a meeting dedicated to the study’s refinement, very minor
changes (for clarity) were made to the rating form, and items relating to the types of knowledge-based material to discuss with F/Os were reviewed and modified slightly. As a small-scale field study, this early exercise allowed limited refinement of the process and verification of the procedures. The exercise also verified the consistency of rater agreement necessary for a F/O performance study. Finally, the short field study also allowed completion of the audit within the 30-day time frame and the budget.

**Instrument**

The “First Officer Evaluation Form” (as reproduced in the Appendix) is very similar to the “Line Evaluation” form used by the same air carrier, and familiar to all instructors, line check airmen and observers. There are eight sections: (a) trip planning; (b) pre-takeoff; (c) takeoff and departure; (d) climb, enroute, and descent; (e) approach and landing; (f) post landing; (g) general; and (h) command ability/CRM. Each of these sections had a number of specific areas to be graded.

Of the total of 74 entries that were graded, 26 were considered “knowledge,” 38 were considered “skills,” and 9 were “abilities.” The 74th entry was for an “overall” score, which was used as a reference to record the observer’s overall evaluation versus a graded, or averaged, score for the same flight. The division of entries into one of the three KSA areas was intuitive to this researcher; any item that required acquiring a specific knowledge of a subject, as taught by the air carrier, was considered “knowledge.” An item that required demonstration of a particular learned physical skill, such as manipulation of flight controls, was considered “skills.” Any personal attribute brought into the air carrier employment environment (not learned) was considered an “ability.”
For those items where a combination of attributes were present (i.e. cockpit/cabin briefing, where the knowledge of what the briefing should contain is combined with the skill of delivery) the particular KSA aspect deemed most important to successfully completing the item was selected as primary.

Many of the items on the First Officer Evaluation Form are the same as those on the line check form, and other items were added or deleted only after a thorough review of the issues and topics necessary to successfully complete the audit. To assure face validity, this task was completed by four SMEs: (a) the Corporate Vice President of Safety; (b) the Director of Flight Safety; and (c) the two check airmen (observers) of the subject airline. Since the observations were to be made in the “real-world” environment, predictive validity was assumed to be very high. A complete breakdown of each section can be found in Table 3, Chapter IV. The list was deemed to be comprehensive by the inclusion of a varied selection of items from each of the three areas to be studied: F/O knowledge, F/O skills, and F/O abilities.

Seven specific items were later decided to be of little use to the PPA, due to a lack of observation during routine flight operations and resultant missing values. Those omitted were (a) international planning, (b) rejected takeoffs, (c) international navigation, (d) holding patterns, (e) monitored approach procedures, (f) non-precision approaches, and (g) autoflight approach procedures. These items exist on the rating form, were graded during a few of the observational flights, and can be found in the final database. However, they have not been included in any of the calculations upon which the final data analysis has been based.
**Likert Scale**

Grading of F/O performance was designed to be quantitative, utilizing the five-point Likert scale. What each observer was looking for, however, was *standardization*. A "1" indicated a performance well below standard; a "5" indicated a performance well above standard. A grade of "3" represented a "standard" F/O, that is, one that performed his/her duties as trained. If rated a 3, the F/O's knowledge base was that required by the air carrier (to a level expected from the training given). It was not expected that many 1 or 5 grades would be recorded; results were in accordance with the expectations.

The referent in this case was the absolute "standard" as expected by the air carrier. Grading to the referent is much more consistent than grading to an "average" performance level. The extensive line check experience of the observers enabled consistent grading to the standard on a continuous, interval scale, throughout the PPA.
CHAPTER IV

ANALYSIS OF DATA

After 30 days of observational flights, the final quantitative data set totaled 64 cases, each case consisting of 6 independent variables and 74 dependent variables. A database was constructed using the Statistical Package for the Social Sciences (SPSS 10.1.0) on a personal computer. Five of the independent variables – years of service, age, total flight time, total airline flight time, and total time in current position – are SPSS scale variables. The sixth variable, aircraft type, is an SPSS nominal variable. All data analysis was completed with the SPSS.

Due to the nature of the study, not all graded events occurred during each one of the observational flights, resulting in missing values in the final data set. This researcher decided that at least 15 observations were required for valid analysis; of the 74 entries, 7 were discarded due to the inability of the observers to gather the required number of observations. The remaining 67 entries were separated into appropriate categories of knowledge, skills, and abilities (see Chapter III for a complete discussion). Table 3, below, subdivides each of the individual variables into the appropriate KSA category. The complete and final version of the grading form used by the observers can be found in the Appendix to this report.
### Table 3

**Rating Form Graded Entries: Knowledge, Skills, and Abilities**

<table>
<thead>
<tr>
<th><strong>Knowledge</strong></th>
<th>Weather review</th>
<th>Computer flight plan</th>
<th>Dispatch procedures</th>
</tr>
</thead>
<tbody>
<tr>
<td>Weight &amp; balance</td>
<td>MEL</td>
<td></td>
<td>Optimum cruise speeds / altitudes</td>
</tr>
<tr>
<td>Weather radar</td>
<td>Fuel management</td>
<td></td>
<td>Diversion / alternate</td>
</tr>
<tr>
<td>FMS / enroute navigation</td>
<td>Descent planning</td>
<td></td>
<td>Speed / altitude restrictions</td>
</tr>
<tr>
<td>Knowledge of minimums / ops spex</td>
<td>Approach planning / briefing</td>
<td></td>
<td>Landing charts</td>
</tr>
<tr>
<td>Terminal speeds / altitudes</td>
<td>Approach nav setup</td>
<td></td>
<td>Use of autoflight systems</td>
</tr>
<tr>
<td>Logbook, maintenance</td>
<td>Equipment knowledge</td>
<td></td>
<td>FOPM knowledge</td>
</tr>
<tr>
<td>Procedures, policies</td>
<td>Company communications</td>
<td></td>
<td>Cockpit discipline</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Skills</strong></th>
<th>Cockpit / cabin briefing</th>
<th>Walkaround</th>
<th>Preflight (cockpit)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ATIS, clearance</td>
<td>Departure nav setup</td>
<td>Engine start, powerback</td>
<td></td>
</tr>
<tr>
<td>Ramp congestion</td>
<td>Engine out taxi</td>
<td>Taxi procedures (out)</td>
<td></td>
</tr>
<tr>
<td>Takeoff performance</td>
<td>Takeoff briefing</td>
<td>Takeoff alert</td>
<td></td>
</tr>
<tr>
<td>Normal takeoff</td>
<td>Crosswind takeoff</td>
<td>Noise abatement procedures</td>
<td></td>
</tr>
<tr>
<td>Minimum speeds, bank angles</td>
<td>SID, departure procedures</td>
<td>Transition level / altitude</td>
<td></td>
</tr>
<tr>
<td>Climb &amp; descent speed control</td>
<td>Autopilot use</td>
<td>Precision / manual approach</td>
<td></td>
</tr>
<tr>
<td>Speed control / stabilized approach</td>
<td>Slot / line up recognition</td>
<td>Go around readiness</td>
<td></td>
</tr>
<tr>
<td>Threshold clearance height</td>
<td>Normal landing</td>
<td>Crosswind landing</td>
<td></td>
</tr>
<tr>
<td>Use of spoilers, reverse, brakes</td>
<td>Taxi procedures (in)</td>
<td>Engine out taxi</td>
<td></td>
</tr>
<tr>
<td>Secure cockpit check</td>
<td>Smoothness, precision of flight</td>
<td>Checklist usage</td>
<td></td>
</tr>
<tr>
<td>Traffic lookout</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Abilities</strong></th>
<th>Attitude</th>
<th>Safety awareness</th>
<th>Professionalism</th>
</tr>
</thead>
<tbody>
<tr>
<td>Personal appearance</td>
<td>Situational awareness</td>
<td>Workload management</td>
<td></td>
</tr>
<tr>
<td>Decision making</td>
<td>Cockpit discipline / sterile cockpit</td>
<td>CRM</td>
<td></td>
</tr>
</tbody>
</table>

Demographics

The 64 cases represent 64 observational flights, but only 61 individual F/Os. Two F/Os were observed once by each observer, and one F/O was observed twice by the same observer. The mean KSA scores and total mean scores for the graded performance of the first two F/Os are shown in Table 4. Note the very close correlation of scores as recorded by the two observers during these four observational flights. Descriptive statistics of all of the independent variables for the complete study are shown in Table 5. The sample size of 64 provided sufficient statistical power for this study.

Table 4

Mean KSA Scores for Duplicated Observations of Two F/Os.

<table>
<thead>
<tr>
<th></th>
<th>Mean knowledge</th>
<th>Mean skills</th>
<th>Mean abilities</th>
<th>Total mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>F/O A</td>
<td>Observer #1</td>
<td>2.90</td>
<td>3.00</td>
<td>2.75</td>
</tr>
<tr>
<td></td>
<td>Observer #2</td>
<td>2.91</td>
<td>2.97</td>
<td>2.75</td>
</tr>
<tr>
<td>F/O B</td>
<td>Observer #1</td>
<td>2.95</td>
<td>2.94</td>
<td>2.63</td>
</tr>
<tr>
<td></td>
<td>Observer #2</td>
<td>3.00</td>
<td>2.94</td>
<td>2.63</td>
</tr>
</tbody>
</table>

Table 5

Descriptive Statistics of Independent Variables

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>Minimum</th>
<th>Maximum</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Years of service</td>
<td>64</td>
<td>.31</td>
<td>12.65</td>
<td>4.15</td>
<td>3.50</td>
</tr>
<tr>
<td>Age (years)</td>
<td>64</td>
<td>28.00</td>
<td>59.00</td>
<td>37.31</td>
<td>6.33</td>
</tr>
<tr>
<td>Total flight time (hours)</td>
<td>62</td>
<td>2424.00</td>
<td>12023.00</td>
<td>6372.10</td>
<td>2373.74</td>
</tr>
<tr>
<td>Airline flight time</td>
<td>64</td>
<td>79.00</td>
<td>8061.00</td>
<td>2312.55</td>
<td>1808.15</td>
</tr>
<tr>
<td>(hours)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total flight time in</td>
<td>64</td>
<td>38.00</td>
<td>2943.00</td>
<td>1003.13</td>
<td>574.94</td>
</tr>
<tr>
<td>current position</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
It should be noted that of the five independent, numeric variables, only “Total time” approaches a normal distribution. This is to be expected, as it is the only variable that includes data brought from outside the immediate environment (flying time recorded prior to being hired by the airline). Plotting the data for the remaining four independent variables resulted in positively skewed distributions, for the following reasons:

1. Airline pilot hiring tendencies are toward younger applicants (age), but occasionally older pilots are accepted for initial employment.
2. F/Os usually, but not always, upgrade to Captain when the opportunity exists (years of service).
3. Overall flight experience within the airline depends almost entirely upon years of service (total airline flight time).
4. Length of service on a particular aircraft is dependent on the staffing requirements of the airline, the working conditions preferred by an individual pilot, and the phasing in and out of various fleet types by the operator (total time in current position).

A survey of the entire F/O population of an air carrier would probably not yield “normal” distributions for the five independent variables. This study’s sample, however, was representative of the subject air carriers F/Os.

Rating Form Entries

All 64 cases were assigned an “overall” score by each observer. To validate this score (the observer’s general impression of the flight), all scores for each individual case were averaged, and then compared to the overall score. Additionally, all scores
considered "knowledge", those considered "skills", and those considered "abilities" were averaged as well. The results are shown in Table 6.

Table 6

*Descriptive Statistics, Overall Scores and Knowledge, Skills, and Abilities Scores*

<table>
<thead>
<tr>
<th>Category</th>
<th>N</th>
<th>Minimum</th>
<th>Maximum</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Observer's overall score</td>
<td>64</td>
<td>2.00</td>
<td>5.00</td>
<td>3.08</td>
<td>.48</td>
</tr>
<tr>
<td>Total mean score</td>
<td>64</td>
<td>2.70</td>
<td>3.24</td>
<td>2.96</td>
<td>.12</td>
</tr>
<tr>
<td>All knowledge items</td>
<td>1243</td>
<td>1.00</td>
<td>5.00</td>
<td>2.92</td>
<td>.38</td>
</tr>
<tr>
<td>All skills items</td>
<td>1654</td>
<td>2.00</td>
<td>5.00</td>
<td>2.98</td>
<td>.28</td>
</tr>
<tr>
<td>All abilities items</td>
<td>493</td>
<td>2.00</td>
<td>5.00</td>
<td>3.15</td>
<td>.55</td>
</tr>
</tbody>
</table>

The range of “observer's overall” scores is high (with a corresponding high standard deviation, or SD) and its mean (3.08) is somewhat higher than the “total mean” score (2.96). The inference is that the observer's overall impression of a flight was somewhat better than the sum of the scores on individual items would indicate. Plotting either the observer's overall score or total mean score against any of the independent variables produced a flat linear regression line for all cases, indicating no trend in overall scores over time-based independent variables.

As can be seen by the small SD values, the mean scores for each category (KSA) are tightly grouped around the mean. “All abilities” has the largest range, due to the influence of individual personalities and the difficulty of implementing successful corporate standardization of attitude, professionalism, and personal appearance. “All skills” has the lowest range and SD, probably due to the emphasis on flying maneuvers/skills as taught by the airline and as practiced daily by line pilots. Figures 1 and 2 plot the mean scores for all categories and the various distributions of each is
Figure 1. All Mean Scores (KSA)/Years of service.

Figure 2. All Mean Scores (KSA)/Total airline flight time.
evident. Note also the linear regressions, which will be considered shortly; knowledge is basically flat, or declines slightly over time; skills basically rises slightly over time, and abilities declines noticeably over time.

Each individual knowledge, skills, and abilities item was extracted from the overall case data and descriptive statistics generated. Table 7 displays the results for all of the knowledge items, sorted in ascending order of means for easier recognition of those areas found to be below standard. Tables 8 and 9 contain the same results for the skills and abilities items, respectively.

For the safety professional, the areas of highest interest are those that encompass the broadest area of knowledge, have a high number of observations (N), and are located near the top or bottom of the table (corresponding to either a high or low relative mean score). Low scores can possibly indicate poor hiring practices, or inadequate or inappropriate training; high scores may serve as models for desirable training techniques in other areas.

Items of specific interest in each category (KSA) were analyzed for possible relationships with the five numeric independent variables (years of service, age, total flight time, airline flight time, and flight time in current position). The particular items chosen were:

1. Knowledge – procedures/policies, equipment knowledge, and cockpit discipline: These three items are ideally representative of the types of knowledge taught at the air carrier; they are located at the top, middle, and bottom of the descriptive statistics table.

2. Skills – use of reverse thrust/brakes, crosswind landing, and normal landing:
Table 7

Descriptive Statistics, Knowledge Items

<table>
<thead>
<tr>
<th>Item</th>
<th>N</th>
<th>Minimum</th>
<th>Maximum</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Landing charts</td>
<td>39</td>
<td>1.00</td>
<td>5.00</td>
<td>2.51</td>
<td>.85</td>
</tr>
<tr>
<td>Procedures/policies</td>
<td>63</td>
<td>2.00</td>
<td>4.00</td>
<td>2.57</td>
<td>.59</td>
</tr>
<tr>
<td>Diversion/alternate</td>
<td>27</td>
<td>1.00</td>
<td>4.00</td>
<td>2.63</td>
<td>.74</td>
</tr>
<tr>
<td>Flight policy Manual</td>
<td>62</td>
<td>2.00</td>
<td>4.00</td>
<td>2.65</td>
<td>.55</td>
</tr>
<tr>
<td>Co. communications</td>
<td>64</td>
<td>2.00</td>
<td>3.00</td>
<td>2.88</td>
<td>.33</td>
</tr>
<tr>
<td>Weather review</td>
<td>62</td>
<td>2.00</td>
<td>4.00</td>
<td>2.90</td>
<td>.39</td>
</tr>
<tr>
<td>Equipment knowledge</td>
<td>64</td>
<td>2.00</td>
<td>4.00</td>
<td>2.92</td>
<td>.48</td>
</tr>
<tr>
<td>Optimum speeds/altitudes</td>
<td>47</td>
<td>2.00</td>
<td>4.00</td>
<td>2.94</td>
<td>.32</td>
</tr>
<tr>
<td>Ops spex, minimums</td>
<td>59</td>
<td>2.00</td>
<td>4.00</td>
<td>2.95</td>
<td>.39</td>
</tr>
<tr>
<td>Approach plan/brief</td>
<td>48</td>
<td>2.00</td>
<td>4.00</td>
<td>2.96</td>
<td>.29</td>
</tr>
<tr>
<td>Use of autoflight systems</td>
<td>44</td>
<td>2.00</td>
<td>3.00</td>
<td>2.98</td>
<td>.15</td>
</tr>
<tr>
<td>Dispatch procedures</td>
<td>63</td>
<td>2.00</td>
<td>3.00</td>
<td>2.98</td>
<td>.13</td>
</tr>
<tr>
<td>Fuel management</td>
<td>46</td>
<td>3.00</td>
<td>3.00</td>
<td>3.00</td>
<td>.00</td>
</tr>
<tr>
<td>Descent planning</td>
<td>46</td>
<td>3.00</td>
<td>3.00</td>
<td>3.00</td>
<td>.00</td>
</tr>
<tr>
<td>Speed/alt restrictions</td>
<td>45</td>
<td>3.00</td>
<td>3.00</td>
<td>3.00</td>
<td>.00</td>
</tr>
<tr>
<td>Terminal speeds/altitudes</td>
<td>45</td>
<td>3.00</td>
<td>3.00</td>
<td>3.00</td>
<td>.00</td>
</tr>
<tr>
<td>Approach nav setup</td>
<td>47</td>
<td>3.00</td>
<td>3.00</td>
<td>3.00</td>
<td>.00</td>
</tr>
<tr>
<td>Logbook, maintenance</td>
<td>64</td>
<td>3.00</td>
<td>4.00</td>
<td>3.02</td>
<td>.13</td>
</tr>
<tr>
<td>Weight &amp; balance</td>
<td>61</td>
<td>3.00</td>
<td>4.00</td>
<td>3.02</td>
<td>.13</td>
</tr>
<tr>
<td>Weather radar</td>
<td>46</td>
<td>3.00</td>
<td>4.00</td>
<td>3.02</td>
<td>.15</td>
</tr>
<tr>
<td>FMS/enroute navigation</td>
<td>45</td>
<td>3.00</td>
<td>4.00</td>
<td>3.02</td>
<td>.15</td>
</tr>
<tr>
<td>Cockpit discipline</td>
<td>64</td>
<td>2.00</td>
<td>4.00</td>
<td>3.03</td>
<td>.53</td>
</tr>
<tr>
<td>Computer flight plan</td>
<td>62</td>
<td>3.00</td>
<td>4.00</td>
<td>3.03</td>
<td>.18</td>
</tr>
<tr>
<td>MEL</td>
<td>34</td>
<td>3.00</td>
<td>4.00</td>
<td>3.06</td>
<td>.24</td>
</tr>
</tbody>
</table>
Table 8

**Descriptive Statistics, Skills Items**

<table>
<thead>
<tr>
<th>Item</th>
<th>N</th>
<th>Minimum</th>
<th>Maximum</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cockpit/Cabin briefing</td>
<td>50</td>
<td>2.00</td>
<td>4.00</td>
<td>2.74</td>
<td>.53</td>
</tr>
<tr>
<td>Traffic lookout</td>
<td>64</td>
<td>2.00</td>
<td>5.00</td>
<td>2.75</td>
<td>.71</td>
</tr>
<tr>
<td>Takeoff briefing</td>
<td>50</td>
<td>2.00</td>
<td>5.00</td>
<td>2.82</td>
<td>.60</td>
</tr>
<tr>
<td>Checklist usage</td>
<td>64</td>
<td>2.00</td>
<td>3.00</td>
<td>2.84</td>
<td>.37</td>
</tr>
<tr>
<td>Transition level/alt</td>
<td>61</td>
<td>2.00</td>
<td>3.00</td>
<td>2.89</td>
<td>.32</td>
</tr>
<tr>
<td>Use of spoilers, rev, brakes</td>
<td>43</td>
<td>2.00</td>
<td>4.00</td>
<td>2.93</td>
<td>.34</td>
</tr>
<tr>
<td>Autopilot use</td>
<td>47</td>
<td>2.00</td>
<td>3.00</td>
<td>2.98</td>
<td>.15</td>
</tr>
<tr>
<td>Takeoff alert</td>
<td>61</td>
<td>2.00</td>
<td>3.00</td>
<td>2.98</td>
<td>.13</td>
</tr>
<tr>
<td>Takeoff performance</td>
<td>61</td>
<td>2.00</td>
<td>4.00</td>
<td>3.00</td>
<td>.18</td>
</tr>
<tr>
<td>Speed, sink, stab. app.</td>
<td>45</td>
<td>2.00</td>
<td>4.00</td>
<td>3.00</td>
<td>.43</td>
</tr>
<tr>
<td>Smoothness precision</td>
<td>39</td>
<td>2.00</td>
<td>4.00</td>
<td>3.00</td>
<td>.32</td>
</tr>
<tr>
<td>Departure nav setup</td>
<td>50</td>
<td>3.00</td>
<td>3.00</td>
<td>3.00</td>
<td>.00</td>
</tr>
<tr>
<td>Engine start, powerback</td>
<td>56</td>
<td>3.00</td>
<td>3.00</td>
<td>3.00</td>
<td>.00</td>
</tr>
<tr>
<td>Engine out taxi</td>
<td>45</td>
<td>3.00</td>
<td>3.00</td>
<td>3.00</td>
<td>.00</td>
</tr>
<tr>
<td>Taxi procedures</td>
<td>40</td>
<td>3.00</td>
<td>3.00</td>
<td>3.00</td>
<td>.00</td>
</tr>
<tr>
<td>Noise abatement proc.</td>
<td>46</td>
<td>3.00</td>
<td>3.00</td>
<td>3.00</td>
<td>.00</td>
</tr>
<tr>
<td>Min speeds, bank angles</td>
<td>46</td>
<td>3.00</td>
<td>3.00</td>
<td>3.00</td>
<td>.00</td>
</tr>
<tr>
<td>SIDs, depart proc.</td>
<td>46</td>
<td>3.00</td>
<td>3.00</td>
<td>3.00</td>
<td>.00</td>
</tr>
<tr>
<td>Climb &amp; descent speeds</td>
<td>45</td>
<td>3.00</td>
<td>3.00</td>
<td>3.00</td>
<td>.00</td>
</tr>
<tr>
<td>Precision / manual app</td>
<td>46</td>
<td>2.00</td>
<td>4.00</td>
<td>3.00</td>
<td>.21</td>
</tr>
<tr>
<td>Threshold clearance height</td>
<td>45</td>
<td>3.00</td>
<td>3.00</td>
<td>3.00</td>
<td>.00</td>
</tr>
<tr>
<td>Taxi procedures</td>
<td>39</td>
<td>3.00</td>
<td>3.00</td>
<td>3.00</td>
<td>.00</td>
</tr>
<tr>
<td>Engine out taxi</td>
<td>42</td>
<td>3.00</td>
<td>3.00</td>
<td>3.00</td>
<td>.00</td>
</tr>
<tr>
<td>Secure cockpit</td>
<td>64</td>
<td>3.00</td>
<td>3.00</td>
<td>3.00</td>
<td>.00</td>
</tr>
<tr>
<td>ATIS, clearance, etc.</td>
<td>64</td>
<td>3.00</td>
<td>4.00</td>
<td>3.02</td>
<td>.13</td>
</tr>
<tr>
<td>Ramp congestion</td>
<td>58</td>
<td>3.00</td>
<td>4.00</td>
<td>3.02</td>
<td>.13</td>
</tr>
<tr>
<td>Slot / line up recognition</td>
<td>46</td>
<td>3.00</td>
<td>4.00</td>
<td>3.02</td>
<td>.15</td>
</tr>
<tr>
<td>Go-around readiness</td>
<td>46</td>
<td>2.00</td>
<td>4.00</td>
<td>3.02</td>
<td>.26</td>
</tr>
<tr>
<td>Preflight (cockpit)</td>
<td>63</td>
<td>3.00</td>
<td>4.00</td>
<td>3.03</td>
<td>.18</td>
</tr>
<tr>
<td>Normal takeoff</td>
<td>46</td>
<td>3.00</td>
<td>4.00</td>
<td>3.04</td>
<td>.21</td>
</tr>
<tr>
<td>Crosswind landing</td>
<td>17</td>
<td>3.00</td>
<td>4.00</td>
<td>3.06</td>
<td>.24</td>
</tr>
<tr>
<td>Walkaround</td>
<td>59</td>
<td>3.00</td>
<td>4.00</td>
<td>3.08</td>
<td>.28</td>
</tr>
<tr>
<td>Crosswind takeoff</td>
<td>19</td>
<td>3.00</td>
<td>4.00</td>
<td>3.11</td>
<td>.32</td>
</tr>
<tr>
<td>Normal landing</td>
<td>45</td>
<td>3.00</td>
<td>4.00</td>
<td>3.11</td>
<td>.32</td>
</tr>
</tbody>
</table>
There are several other items located at the top of the skills ascending means table (e.g. briefings, traffic lookout) but the items selected are more typical of the types of physical flying skills of greater interest in this study.

3. Abilities – CRM, attitude, professionalism: CRM was selected because of the emphasis placed on it during F/O initial and recurrent training, while attitude and professionalism both scored high, had high SDs and are typical of the abilities of interest to the PPA.

Knowledge

The “all knowledge items” mean score of 2.92 is markedly below standard. Each case mean knowledge score was plotted as a linear regression against the five independent numeric variables. (A linear curve estimation model was selected for simplicity). The results are similar in each case. As illustrated in Figures 1 and 2, the mean knowledge trend lines start below standard and display either no improvement, or
decrease as a function of both years of service and total airline flight time. The start of the regression line (below standard) and its lack of rise over either years of service or total airline flight time indicates that the knowledge-based material was not taught adequately in initial training and has not been learned during subsequent time with the air carrier.

The items chosen for additional scrutiny – policies/procedures, equipment knowledge, and cockpit discipline – are illustrated in Figures 3 and 4, relative to years of service and flight time in current position. The regression lines for these items are typical of most knowledge items – starting at or slightly below standard, and either remaining flat or declining slightly over time. These particular items reinforce the trend lines discovered previously in the knowledge mean scores.

Of the 24 knowledge items, 40% of the resultant regression lines start at a point below standard. All trend lines decrease over time, to varying degrees. Interestingly, however, a few variables demonstrate tendencies that do not follow this norm. The regression lines for both equipment knowledge and Flight Policy Manual knowledge rise slightly when plotted against age. These plots do not correlate with the rest of the data, and therefore may be insignificant. However, one possible explanation for this atypical performance may be the increasing maturity of the subjects (with age) and subsequent study habits/motivation to learn.

**Skills**

The “all skills items” mean score of 2.95 is slightly below standard. Each case mean skills score was compared to the five independent numeric variables considered
Figure 3. Selected knowledge items/Years of service

Figure 4. Selected knowledge items/Flight time in current position
previously, and a linear regression generated. The results are virtually identical – the regression line starts slightly below standard and rises to standard or above in all cases. The implication for skills is that initial training is good, producing a product very close to the standard. As one would expect, gaining flying experience with the air carrier increases the skill level of the pilot, either in terms of years of service or in time in current position.

The three areas of skills interests – use of reverse thrust/brakes, crosswind landing, and normal landing – are diagramed in Figures 5 and 6, plotted against years of service and total time in current position. They are typical of most of the skills graded events; the regression line in both charts follows the general pattern established by the mean skills scores, starting slightly below, or at standard, and rising over time.

Abilities

The “all abilities” mean score of 3.15 is well above standard. Each case mean abilities score was compared to the five independent numeric variables considered previously, and a linear regression performed. Four of the five (years in service, age, total time, and total airline flight time) are virtually identical, with regression lines that start well above standard and decrease over time. One (time in current position), however, is almost flat.

The implications raised by the descending trend lines are serious; the personal attributes of pilots when hired (attitude, professionalism, personal appearance, etc.) are scored as very high, but decrease markedly over time, whether measured as years of service or total airline flight time.
Figure 5. Selected skills items/Years of service

Figure 6. Selected skills items/Flight time in current position
Figures 7 and 8 illustrate the abilities areas of interest – CRM, attitude, and Professionalism – plotted against years of service and total airline flight time. Note that while all scores start above standard, all decrease over time to a point at or below standard, in a manner very similar to the abilities mean scores previously diagrammed.

Probationary Pilot Scores

Traditionally, air carriers use the first year of pilot employment as a probationary period. During that time, probationary pilot performance is continually monitored by management personnel, and a line operational checkride must be completed successfully prior to the pilot being offered permanent employment status.

Two important pre-employment considerations to an air carrier are age (years of service remaining to the carrier) and total flight time (experience brought to the job). Of particular interest to this study, therefore, are probationary pilot’s mean scores and total mean scores in relation to age and total flight time. A total of 11 probationary F/Os were observed during the PPA; the descriptive statistics are contained in Table 10.

Comparing the total mean and KSA mean scores to those of the overall PPA study (Table 6) shows similar, though not identical mean scores. Interestingly, however, the SDs for all three KSA categories are much lower (.07 vs. .38, .14 vs. .28, and .26 vs. .55). The lower SDs could be the result of a smaller sample group, or it could be that the variations in pilot performance are much less after initial training than found in the total study group.

Figures 9 and 10 chart the same mean values (KSAs) against the independent variables of age and total flight time. Linear regressions were not used in these cases; a
Figure 7. Selected abilities items/Years of service

Figure 8. Selected abilities items/Total airline flight time
Figure 9. Probationary pilot mean scores/Age

Figure 10. Probationary pilot mean scores/Total flight time
Table 10

*Descriptive Statistics, Probationary Pilots Mean Scores*

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>Minimum</th>
<th>Maximum</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean knowledge</td>
<td>11</td>
<td>2.83</td>
<td>3.05</td>
<td>2.93</td>
<td>.07</td>
</tr>
<tr>
<td>Mean skills</td>
<td>11</td>
<td>2.70</td>
<td>3.16</td>
<td>2.91</td>
<td>.14</td>
</tr>
<tr>
<td>Mean abilities</td>
<td>11</td>
<td>2.88</td>
<td>3.57</td>
<td>3.24</td>
<td>.26</td>
</tr>
<tr>
<td>Total mean</td>
<td>11</td>
<td>2.81</td>
<td>3.16</td>
<td>2.98</td>
<td>.11</td>
</tr>
</tbody>
</table>

quadratic curve estimation model, while more complex, provides a better understanding of the overall trend. Assuming that mean performance scores aren’t measurably affected during the first year of service with an air carrier, the data suggests that there may be performance patterns based substantially on qualities brought into the work environment. Specifically, all three categories (KSAs) trend upward with age (i.e., the older the new hire pilot is, the better he/she may be expected to perform during his/her first year). Additionally, all three categories (KSAs) tend to reach a peak in performance at about 7,000 hours of flight time (perhaps indicating an “ideal” total flight time for new hire pilots).

The sample size in this particular analysis (N = 11) is very small, and no attempt has been made to determine if there is a significant difference between the means at a given probability level (t test or ANOVA). Thus, the level of confidence in any specific implication or regression may not be high; however, in general, the trends may indicate an area of additional study that could reap significant benefits for air carriers in the future.
Fleet-wide Studies

The analysis of one variable related specifically to the type of aircraft operated; “traffic lookout.” It is well known that automation in the aircraft cockpit is intended to reduce pilot workload, thus providing more time for human-oriented tasks such as traffic watch. Industry experience, however, suggests that the additional machine monitoring duties imposed on the pilot may negate any perceived improvement in overall workload.

There are basically four types of aircraft in the subject air carrier’s fleet:

1. Fleet type 1 – the original “round-dial,” analog cockpits of the early 1970s (e.g., DC 9-30, B 727-200).
2. Fleet type 2 – transition cockpits consisting of capable autopilots, Flight Management Systems (FMSs) and autothrottles (e.g., MD-80, B 737-300).
3. Fleet type 3 – mostly-glass cockpits with a complete FMS, moving map displays, autothrottles, and performance enhancement computers (e.g., B 757-200, B 767-200).
4. Fleet type 4 – all glass cockpits with the most up-to-date navigation and aircraft control systems available (e.g., B 717, B 777).

Figure 11 diagrams the results of the PPA regarding traffic lookout. As automation incrementally increases (Fleet 1 through Fleet 4) the mean score for traffic lookout improves as well, suggesting that automation does provide workload benefits that may not be intuitively apparent to the industry.

Table 11 documents the comparison between all areas (KSA) and the various fleet types. Note the consistent rise in scores from the least automated to most automated
Figure 11. Traffic lookout mean scores/Fleet type

The only variable that does not follow the trend is the high mean score for knowledge (Fleet 1) which is higher than Fleets 2 and 3, and higher than would be expected relative to the

Table 11. Mean Knowledge, Skills and Abilities for Each Fleet Type.

<table>
<thead>
<tr>
<th>Fleet Type</th>
<th>N</th>
<th>Mean knowledge</th>
<th>Mean skills</th>
<th>Mean abilities</th>
<th>Total mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fleet 1</td>
<td>3</td>
<td>2.91</td>
<td>2.80</td>
<td>2.96</td>
<td>2.88</td>
</tr>
<tr>
<td>Fleet 2</td>
<td>30</td>
<td>2.90</td>
<td>2.93</td>
<td>3.15</td>
<td>2.96</td>
</tr>
<tr>
<td>Fleet 3</td>
<td>24</td>
<td>2.90</td>
<td>2.98</td>
<td>3.15</td>
<td>2.97</td>
</tr>
<tr>
<td>Fleet 4</td>
<td>7</td>
<td>2.93</td>
<td>2.96</td>
<td>3.26</td>
<td>2.99</td>
</tr>
</tbody>
</table>
other variables. Interestingly, Fleets 2, 3, and 4 utilize computer-based instruction for many portions of initial training; Fleet 1 is the only aircraft type that employs personal instructors exclusively for all “knowledge” based training.

The skills variable score for Fleet 1 appears slightly lower than would be expected. It is possible that since the most junior pilots were usually assigned to that aircraft, those F/Os had not yet had the opportunity to acquire or perfect their flying skills. It may also be that the automation available in Fleets 2, 3, and 4 masks the fact that the F/Os flying those aircraft did not need to demonstrate the “stick and rudder” skills necessary to operate the aircraft in Fleet 1.
In the air carrier environment, the data generated by this PPA would be of major interest to managers and directors in the safety, training, and flight standards departments. Human resources professionals would want to be involved in the analysis of data that might improve pilot hiring techniques, and aircraft fleet managers would be eager to review specific aircraft/pilot performance information.

Evaluation of the quantitative aspects of the data generated by this PPA will continue in Chapter V. Conclusions and recommendations will be presented, augmented by additional qualitative observations and discussion.

Qualitative Data

An important aspect of this PPA was the unrestricted flow of information between the F/O population and the air carrier's safety department. Additionally, the auditors recorded their personal observations of general safety issues during the course of the PPA. That qualitative data is summarized in this section – not so much for the purpose of scientific analysis, but to present a more comprehensive overview of the operating environment under which the PPA was conducted.

*Personal Observations of Flight Crews*

Once assured of the focus of the audit – anonymous data collection, not specific, individual compliance verification with FARs or air carrier policies and procedures – T/Os were eager to express their opinions to the observers. Those included:

1. **Training** – certain components of basic indoctrination, international procedures, and security courses were regarded as “boring” and “ineffective,”
with several pilots calling it “fill the square” training. Others felt there was too much “training by bulletin,” which also is generally perceived as ineffective.

2. Communication – while unrestricted flow of information between management and crews is important, many F/Os felt that the methods used were too diverse. Suggestions were submitted to coordinate the various electronic (integral, crew-based computer system), paper, and web-based technologies, for easier dissemination to the crews.

3. Operational – ground (vehicular) traffic on the ramp areas is a problem at the major airports, particularly those that are “hubs” for the carrier. Several suggestions were for additional “policing” of ramp traffic. Also, it was noted that flight attendants and gate personnel need reminders not to interrupt flight crews while completing checklists.

Personal Observations of the Auditors

There were many observations recorded on the evaluation forms, most of which were directly related to F/O performance. Those served only to clarify scoring. But in their final reports, the auditors also pointed out several other areas of concern, ones that were not easily categorized on the forms, or related more to the operation in general. They included:

1. Altitude awareness – many F/Os do not fully understand minimum altitudes as presented on Jeppesen approach and enroute charts, and adherence to the air carrier’s own altitude verification procedures are not always complied with. Most F/Os do not have a good working understanding of the new National
Route Plan (NRP) requirements and procedures.

2. CRM – the observers noted a general lack of verbalization between crew members.

3. Crew callouts – a general lack of knowledge regarding several specific callouts, for altitude awareness, altimeter use, and engine monitoring.

4. Captain as “mentor” – interestingly, one auditor observed an F/O on two different occasions, flying with two different Captains. On the first flight, the Captain was a very “laid-back” individual, generally relaxed about procedural compliance. The F/O’s performance was rated as below average. On the second trip, the F/O was teamed with a very professional, more demanding Captain. The F/O’s performance (observed scores) increased dramatically.
CHAPTER V
CONCLUSIONS & RECOMMENDATIONS

“The foundation of air safety is essentially the willingness to examine, recognize, report, and discuss conditions that require remedy” (Air Safety Investigator Quinn, Quantas Airlines, personal communication, 1999). Mr. Quinn’s observation is a valid, and important one. The fundamental assumption in this premise, however, is that the data used as a basis for examination, recognition, reporting, and discussion, is available. It is the knowledge of “conditions that require remedy” that is the first, and primary, prerequisite for an effective air safety program.

The PPA

The research question initially asked was, “Within an air carrier’s pilot population, are their KSAs directly related to the pilot’s age, company longevity, total flight time, or time in current position?” The results for this PPA show that there is a definite relationship between these independent variables and F/O KSAs, and that the resultant safety implications for an air carrier can be significant.

Knowledge

The “all knowledge” items mean score was 2.92, and 40% of the knowledge item’s charted linear regression lines originated at or below standard. These two findings
are indicative of a pilot population that has not been fully educated in those areas examined during the study, or have rapidly lost knowledge gained during training.

Another factor could be that as a new F/O became more comfortable with his professional duties, complacency began to set in. Identifying this pattern and interceding to reverse it is a primary responsibility of the flight safety professional.

The lowest knowledge scores were recorded in landing chart usage, policies and procedures, diversion procedures, flight manuals, and company communications. Other knowledge variables were found to be below standard as well.

The regression plots of virtually all knowledge items over time (either years of service, time in current position, or total airline flight time) were flat or declined slightly, indicating that the F/O’s knowledge of these basic items does not increase once the pilot has started line flying. The independent variable of total flight time was not in itself considered significant in this analysis, as the other “time-related” variables provided the necessary data. Age of F/Os had virtually no bearing on the results, other than when plotted against equipment knowledge and Flight Policy Manual knowledge. As stated previously, it is not clear why those two items do not correlate with all other knowledge data patterns.

Perhaps the most disturbing finding, and the one with the greatest negative trend, was F/Os’ lack of knowledge of certain operational policies and procedures. This is particularly applicable to those operations outside the normal, daily routine. In a few cases, this could be a function of the type of training received, i.e. “training by bulletin.” The pilots themselves commented on their negative perceptions of certain aspects of the air carrier’s training (see Qualitative Data, Chapter IV).
Skills

The “all skills” items mean score was only slightly below standard. Several items, including cockpit/cabin briefings, traffic lookout, checklist usage, and use of spoilers, brakes, reverse thrust, and autopilot, were identified as below standard. However, the linear regression plots for most of the skills items indicated that pilot skills in general improve over time, whether measured against years of service, time in current position, or total airline flight time.

The data indicated that the F/Os are acquiring the necessary skills while in training, and continue to develop them once flying on the line. Skills that rely on physical manipulation of aircraft controls, such as crosswind landing, crosswind takeoff, and normal landing improve at a greater rate than do observed interpersonal skills.

Abilities

The “all abilities” item mean score of 3.15 is well above standard. In particular, areas critical to the safety of flight, such as CRM, decision-making, and safety awareness scored well above standard. Personal appearance, attitude, and professionalism received the highest scores in the audit. One interesting finding was that situational awareness was the only abilities item to score below standard, which is significant because SA has been identified in many studies as being a factor in recent CFIT accidents.

The regression line for each abilities item plotted against any of the time-related variables indicated a strong downward-performing trend. While the personal attributes of the F/Os hired are outstanding, negative influences of line flying markedly decreased virtually all scores.
Probationary Pilot Scores

The mean KSA scores of new-hire pilots (less than 1 year of service) was plotted against age and total flight time, and a quadratic regression performed. Results indicated that the peak age for all KSA mean scores as well as total mean was related to age, with (generally) older being better. Performing the same calculations with total flight time indicated that there is a definite peak time for first-year pilot performance in every category: approximately 7,000 hours of flight time.

Traditionally, air carriers have attempted to hire the youngest pilots possible with the minimum experience necessary. Additional experience was welcomed, but not necessarily at the expense of age. The sample size of the probationary pilot study group in this study was small, so this researcher’s confidence level in conclusions drawn from this data is limited. However, additional research may produce similar results. That is, that the guidelines used for pilot hiring may need to be reevaluated; emphasis on increased age and “ideal” experience levels may be of significant benefit to an air carrier.

Fleet-wide Studies

This study has generated some interesting observations. First, the level of automation utilized in an aircraft cockpit related directly to the amount of outside traffic watch practiced by the F/O (more automation allows for more outside scan). Future studies could focus on this particular issue, thus reinforcing the argument that automation, when properly understood and monitored, does provide a greater level of flight safety, in terms of the traffic separation environment.
Secondly, while the total mean scores were directly related to the aircraft type, with the lowest scores on the oldest, least-automated aircraft, the relative mean knowledge scores were the highest for the oldest aircraft. Data from future studies might clarify whether this phenomenon derives from two slightly different F/O populations (e.g., less experienced vs. older), or from differences in training, such as human vs. computer-based instruction methods.

Finally, the fleet-wide data highlighted an unexpected disparity in flying skills. The older, less automated aircraft produced the lowest skills scores, while the most automated produced the highest scores. Perhaps F/Os flying the automated aircraft actually did demonstrate a higher skill level; or, because the aircraft was “flying itself” as a result of more advanced technology, the observers might only have had the impression of higher skill levels. Conversely, the non-automated aircraft F/O, flying the aircraft manually, might have demonstrated what appeared to be a lower level of skill, but might actually be more capable in the traditional “flight skills” sense. Future research could be designed to clarify this issue.

Air Safety Programs

The second research question asked at the beginning of this study was: “Could an individual pilot’s KSA data be integrated into the air carrier’s overall data collection process to enhance flight safety?” The simple answer is, “Absolutely!”

For flight operations, it is essential that the flight performance data gathered is both accurate and appropriate. Safety programs have traditionally focused their efforts in the following areas:
1. FOQA provides excellent data on the performance of an aircraft (and to a lesser extent the pilot) in the line environment. The information is most useful for a general overview of flight operations events and exceedances, or review of specific “incident” flights: Strengths – identification of aircraft and flight crew operational performance; weaknesses – negative performance monitoring only, no specific individual pilot knowledge or abilities rating.

2. ASAP identifies operational safety issues by thoroughly documenting potential pilot violations, including deviations from clearances and established procedures: Strengths – volume of data allows identification and continual monitoring and trending of safety issues; weaknesses – negative performance monitoring only, limited individual pilot skills information, no pilot knowledge or abilities ratings.

3. LOSA provides a crew-based perspective of line operations: Strengths – thorough assessment of crew-based performance in the line environment; weaknesses – very limited individual pilot information observed or recorded. The LOSA is complementary to a PPA, but its primary purpose is exploration of human performance related to threat and error management, rather than pilot performance measurement.

4. AQP provides both individual and crew-based information on pilot skills in a “rehearsed” environment: Strengths – accurate observations and ratings of pilot performance during specific simulator sessions and LOFT scenarios; weaknesses – provides no information regarding pilot knowledge, skills, or abilities operating in the actual line environment.
Specifically, quantitative line pilot performance data that is difficult to collect in traditional safety programs is the primary focus of a PPA. Table 1 in Chapter I of this report highlights those types of data that, for the most part, are generated exclusively by the application of a PPA—individual pilot knowledge, and individual pilot abilities.

All aviation safety professionals would agree that having access to any of this information is important. The air carrier safety experience of this researcher is the basis for the belief that all of it is necessary to build a complete, comprehensive aviation safety program. Each component complements the next; each is necessary to the success of the overall safety process.

Figure 13 is an illustration of how the overall process might function. The loop diagram can be entered at any point, depending on where the initial safety deficiency is identified. For example, FOQA data might indicate that final approaches flown in a particular aircraft type at a specific airport are tend to be “high energy,” and not fully stabilized. ASAP reports would confirm the air carrier’s line pilot’s concerns with these specific approach events. Procedures and policies, if necessary, would be modified, and those changes issued to the flight crews. If possible, air traffic control (ATC) personnel could be included in the process for possible modification of traffic handling techniques. AQP could confirm that the new procedures are being trained properly, and a subsequent PPA would be designed to determine any problems associated with the operational implementation of the policies, thus verifying the augmented training. A LOSA could then confirm, from a system safety perspective, that the process has been effective. FOQA would then be used to monitor approaches at the airport, confirming that the
problem has been resolved. If operational problems persist, the process would begin again, and the loop would continue until consistently stable approaches were observed.

Each individual program is an integral component of the loop process, identifying specific air carrier deficiencies, or monitoring or confirming specific operational issues. But it is also important to note that not every step of the loop is necessary for each safety deficiency being addressed. Frequently, specific issues can be resolved using just a few of the steps in the loop. The full complement of programs, however, assures that every aspect of the air carrier’s operation is fully monitored.

Figure 13. The system safety loop of air carrier safety programs

Recommendations

A thorough review of instructional techniques (computer-based vs. human, teaching aids, etc.) would be beneficial in determining suitability in the current air carrier training environment. Additional emphasis is needed in both initial and recurrent training on those subject areas (knowledge and skills) found by this study to be below standard.
It would be appropriate to review ASAP and FOQA databases to determine if below standard areas (within any KSA group) are the subjects of, or related to, other recent program findings at the subject air carrier. Analysis of those relationships (if any) could determine the most effective way to improve education (knowledge), training (skills), and personal performance (abilities).

As noted in the qualitative data, a F/O’s performance can be dramatically altered by the level of professionalism demonstrated by the Captain. F/Os usually will adopt the Captain’s “culture,” even if it is in psychological disagreement with the F/O’s style. CRM experience confirms that junior F/Os will seek a certain comfort level by not outperforming their job-related superior (Captain). The importance of the role the Captain plays in developing and maintaining a professional culture among F/Os cannot be overstated. Air carriers must recognize this fact and implement programs to encourage Captains to assist in the continual professional growth of the new-hire pilot group.

The use of proven programs, such as experienced peer mentoring and probationary interviews and testing requirements, should be encouraged. These programs can stimulate additional study and understanding of the knowledge-based requirements of the air carrier. The mentors could also provide support for continual development of personal abilities/attributes, e.g. professionalism, personal appearance, etc.) as well as providing a continual conduit of two-way information flow.

Air carrier safety professionals must be alert to tendencies toward complacency within any pilot group, and devise methods to reverse those trends. ASAP and FOQA can be particularly useful in this regard.
Initial and recurrent training requirements should be reassessed, perhaps with an enhanced emphasis on knowledge based issues. The industry has been successful at training flight skills to proficiency, and those skills must continue to be monitored. However, with the increase in the use of flight automation, a fundamental shift in educational thought might be appropriate.

A scientifically designed, professionally implemented PPA should be utilized by any air carrier on a regular basis. The audit should focus on pilots within a particular aircraft fleet type, crew position, experience level, or other specific group within the air carrier's pilot population. The data should be combined with information from other safety programs to determine appropriate corrective measures (if necessary) and to recommend follow-on studies within the overall program. Maintenance of absolute confidentiality is a must; accurate records should be retained for comparative analysis from study to study.

A well-designed, appropriately conducted PPA can be a valuable addition to the air carrier's safety program toolbox. The PPA augments the overall data collection process, confirms findings, and supports the recommendations resulting from collaborative use of the other fundamental safety programs – LOSA, FOQA, ASAP and AQP. As the industry continually changes, particularly with the introduction of “next-generation” concepts like the Small Aircraft Transportation System (SATS), evolution of aviation safety philosophy will be essential. The safety program loop and the PPA are integral to the safe, successful air transportation system of the future.
REFERENCES


APPENDIX

PILOT PROFICIENCY AUDIT FORM
# [COMPANY NAME] CORPORATE SAFETY

Internal Evaluation Program, Study 01-01

<table>
<thead>
<tr>
<th>A/C Type</th>
<th>DoH</th>
<th>PF/PM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>M/F</td>
<td>Prob. Comp.</td>
</tr>
<tr>
<td>Total T</td>
<td>Airline Time</td>
<td>TT in cur.pos</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Trip Planning</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>Approach and Landing</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>International planning</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Knowledge of min, ops spex</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Weather review</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Approach planning / briefing</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Computer flight plan</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Landing charts</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dispatch procedures</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Terminal speeds / altitudes</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Weight &amp; Balance</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Approach nav setup</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MEL</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Use of autoflight systems</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Pre-Takeoff</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>Non-precision app</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Cockpit/Cabin briefing</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Precision / manual app</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Walkaround</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Precision / autoflight app</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Preflight (cockpit)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Speed / sink control, stab app</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ATIS, clearance, etc.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Slot / line up recognition</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Departure nav setup</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Go-around readiness</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Eng start, powerback</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Threshold clearance height</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ramp congestion</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Normal landing</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Engine out taxi</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Crosswind landing</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Taxi procedures</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Use of spoilers, reverse, brks</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T/O performance</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T/O briefing</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T/O alert</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Takeoff and Departure</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>Secure cockpit</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Rejected T/O</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Logbook, maint</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Normal T/O</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>General</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>Crosswind T/O</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Attitude</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Noise abatement procedures</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Safety awareness</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Min man speeds, bank angles</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Flight Ops policy knowledge</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SID, departure procedures</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Flight Ops policy knowledge</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Climb, Enroute, Descent</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>Procedures / policies</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Transition level/alt</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Professionalism</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Climb &amp; descent speeds</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Company communications</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Autopilot use</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Smoothness / precision of flt</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Optimum cruise speeds/alt</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Checklist usage</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Weather – Radar</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Traffic lookout</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fuel management</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Personal appearance</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Int'l navigation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Command Ability / CRM</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Diversion / alternate</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Situational awareness</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FMS / enroute navigation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Workload management</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>Descent planning</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Decision Making</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Speed / altitude restrictions</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Cockpit discipline sterile ckpt</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Holding</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>CRM</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
## Overall Evaluation:

|--------------|-----------------|-------------------|------------|------------------|-------------|

**Comments**