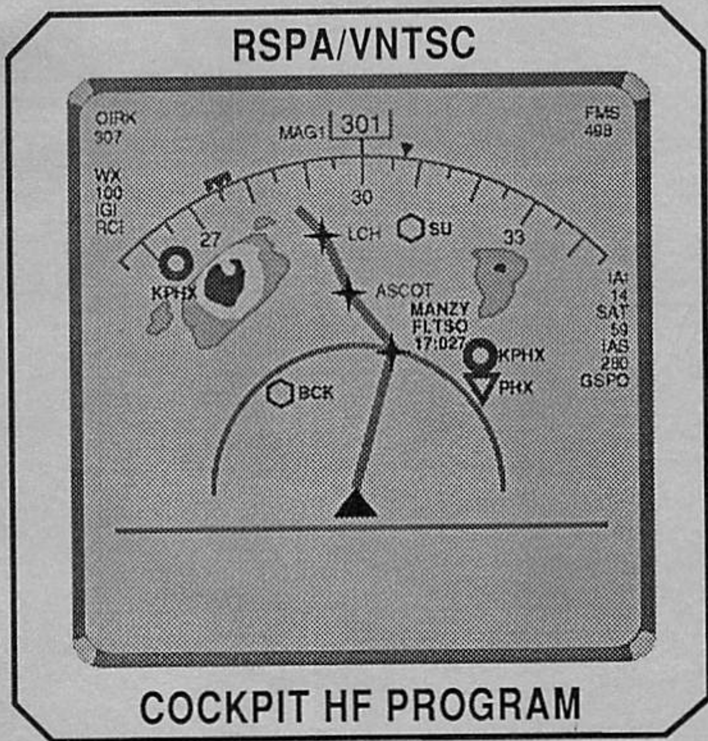
DOT/FAA/RD-93/26
DOT-VNTSC-FAA-93-6
Office of Research
and Development Service
Washington, DC 20591

Integrated Measurement of Crew Resource Management and Technical Flying Skills



George L. Kaemph, Ph.D.
David W. Klinger

Klein Associates, Inc.
502 E. Dayton-Yellow Spring Road
Fairborn, OH 45324-3987

Final Report
August 1993

This document is available to the public
through the National Technical Information Service
Springfield, Virginia 22161

U.S. Department of Transportation
Federal Aviation Administration

REPORT DOCUMENTATION PAGE

Form Approved OMB No. 0704-0188

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	August 1993	Final Report; February 1992-May 1993

4. TITLE AND SUBTITLE
Integrated Measurement of Crew Resource Management and Technical Flying Skills

5. FUNDING NUMBERS
FA2E2/A2155

6. AUTHOR(S)
George L. Kaempf, Ph.D.
David W. Klinger

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
*Battelle Memorial Institute
505 King Avenue
Columbus, OH 43201-6424

**Klein Associates, Inc.
582 E. Dayton
Yellow Spring Road
Fairborn, OH 45324-3987

8. PERFORMING ORGANIZATION REPORT NUMBER
DOT-VNTSC-FAA-93-6

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Department of Transportation
Federal Aviation Administration
Research and Development Service
Washington, DC 20591

10. SPONSORING/MONITORING AGENCY REPORT NUMBER
DOT/FAA/RD-93/26

11. SUPPLEMENTARY NOTES
*Under contract to: U.S. Department of Transportation
Research and Special Programs Administration
John A. Volpe National Transportation Systems Center
Cambridge, MA 02142-1093

**Under contract to: Battelle Memorial Institute

12a. DISTRIBUTION/AVAILABILITY STATEMENT
This document is available to the public through the National Technical Information Service, Springfield, VA 22161

12b. DISTRIBUTION CODE

13. ABSTRACT (Maximum 200 words)

This report presents the findings of a study designed with two objectives: to produce a prototype performance measurement instrument (PMI) that integrates the assessment of Crew Resource Management (CRM) and technical flying skills and to investigate the suitability of the Critical Decision method (CDM) for eliciting expert information concerning performance measurement. The work was funded by the FAA in support of the Advanced Qualification Program (AQP) and conducted in cooperation with a major U.S. carrier. The researchers used CDM to identify critical components of performance assessment for specific flight tasks and developed a prototype PMI. The instrument contains two sections for each task. One section allows an evaluator to record significant pilot and crew behaviors observed; the second section allows the evaluator to provide a subjective assessment of pilot and crew proficiency. The researchers pretested the instrument and made revisions based on recommendations from experienced instructors. The researchers then evaluated the PMI with eight instructors observing a total of 16 different flight crews in recurrent training, performing a standard Line Oriented Flight Training (LOFT) scenario in a flight simulator. The instructors reliably and accurately employed the PMI to assess performance of the crew and the individual pilot. The authors recommended that AQP developers use Cognitive Task Analysis (CTA) techniques to develop training programs for cognitive and team tasks.

14. SUBJECT TERMS
Human Factors, Evaluation, Cockpit Resource Management, CRM, Critical Decision Method, CDM, Line Oriented Flight Training.

15. NUMBER OF PAGES
113

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	

NSN 7540-01-280-5500

Standard Form 298 (Rev. 2-89)
Prescribed by ANSI Std. 239-18
298-102

REPORT DOCUMENTATION PAGE

1. AGENCY USE ONLY (Leave blank)

2. REPORT DATE
August 199[illegible]

3. REPORT TYPE AND DATES COVERED
Final Report [illegible]

4. TITLE AND SUBTITLE
Interrater and Measurement of [illegible] Technical Flying Skills

6. AUTHOR(S)
George L. Kaempf, Ph.D.
David W. Klinger

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Klein Associates Inc.
[illegible] Avenue
[illegible], OH 45[illegible]

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Department of Transportation
Federal Aviation Administration
Research and Development Service
Washington, DC 20591

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT

This document is available [illegible]
Technical Information [illegible]

12b. DISTRIBUTION CODE

13. ABSTRACT (Maximum 200 words)

This report presents the findings [illegible] measurement instrument (PMI) [illegible] to investigate the [illegible] performance [illegible] and [illegible] in [illegible] for [illegible]. One section of the [illegible] the evaluator [illegible] the instrument and [illegible] the PMI with [illegible] standard [illegible] employed the PMI [illegible] cognitive [illegible]

14. SUBJECT TERMS
Human Factors, Evaluation, [illegible]

15. NUMBER OF PAGES

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT
Unclassified

18. SECURITY CLASSIFICATION OF THIS PAGE
Unclassified

19. SECURITY CLASSIFICATION OF ABSTRACT
Unclassified

20. LIMITATION OF ABSTRACT

NSN 7540-01-280-5500

PREFACE

This report was prepared by Klein Associates, under subcontract to Battelle Memorial Institute. It presents the findings of a study designed with two objectives: to produce a prototype performance measurement instrument (PMI) that integrates the assessment of Crew Resource Management (CRM) and technical flying skills and to investigate the suitability of the Critical Decision method (CDM) for eliciting expert information concerning performance measurement. The work was funded by the FAA in support of the Advanced Qualification Program (AQP) and conducted in cooperation with a major U.S. carrier. The researchers used CDM to identify critical components of performance assessment for specific flight tasks and developed a prototype PMI. The instrument contains two sections for each task. One section allows an evaluator to record significant pilot and crew behaviors observed; the second section allows the evaluator to provide a subjective assessment of pilot and crew proficiency. The researchers pretested the instrument and made revisions based on recommendations from experienced instructors. The researchers then evaluated the PMI with eight instructors observing a total of 16 different flight crews in recurrent training, performing a standard Line Oriented Flight Training (LOFT) scenario in a flight simulator. The instructors reliably and accurately employed the PMI to assess performance of the crew and the individual pilot. The authors recommend that AQP developers use Cognitive Task Analysis (CTA) techniques to develop training programs for cognitive and team tasks.

METRIC / ENGLISH CONVERSION FACTORS

ENGLISH TO METRIC

LENGTH (APPROXIMATE)

1 inch (in) = 2.5 centimeters (cm)
1 foot (ft) = 30 centimeters (cm)
1 yard (yd) = 0.9 meter (m)
1 mile (mi) = 1.6 kilometers (km)

AREA (APPROXIMATE)

1 square inch (sq in, in^2) = 6.5 square centimeters (cm^2)
1 square foot (sq ft, ft^2) = 0.09 square meter (m^2)
1 square yard (sq yd, yd^2) = 0.8 square meter (m^2)
1 square mile (sq mi, mi^2) = 2.6 square kilometers (km^2)
1 acre = 0.4 hectares (he) = 4,000 square meters (m^2)

MASS - WEIGHT (APPROXIMATE)

1 ounce (oz) = 28 grams (gr)
1 pound (lb) = .45 kilogram (kg)
1 short ton = 2,000 pounds (lb) = 0.9 tonne (t)

VOLUME (APPROXIMATE)

1 teaspoon (tsp) = 5 milliliters (ml)
1 tablespoon (tbsp) = 15 milliliters (ml)
1 fluid ounce (fl oz) = 30 milliliters (ml)
1 cup (c) = 0.24 liter (l)
1 pint (pt) = 0.47 liter (l)
1 quart (qt) = 0.96 liter (l)
1 gallon (gal) = 3.8 liters (l)
1 cubic foot (cu ft, ft^3) = 0.03 cubic meter (m^3)
1 cubic yard (cu yd, yd^3) = 0.76 cubic meter (m^3)

TEMPERATURE (EXACT)

$[(x - 32)(5/9)]$ °F = y °C

METRIC TO ENGLISH

LENGTH (APPROXIMATE)

1 millimeter (mm) = 0.04 inch (in)
1 centimeter (cm) = 0.4 inch (in)
1 meter (m) = 3.3 feet (ft)
1 meter (m) = 1.1 yards (yd)
1 kilometer (km) = 0.6 mile (mi)

AREA (APPROXIMATE)

1 square centimeter (cm^2) = 0.16 square inch (sq in, in^2)
1 square meter (m^2) = 1.2 square yards (sq yd, yd^2)
1 square kilometer (km^2) = 0.4 square mile (sq mi, mi^2)
1 hectare (he) = 10,000 square meters (m^2) = 2.5 acres

MASS - WEIGHT (APPROXIMATE)

1 gram (gr) = 0.036 ounce (oz)
1 kilogram (kg) = 2.2 pounds (lb)
1 tonne (t) = 1,000 kilograms (kg) = 1.1 short tons

VOLUME (APPROXIMATE)

1 milliliter (ml) = 0.03 fluid ounce (fl oz)
1 liter (l) = 2.1 pints (pt)
1 liter (l) = 1.06 quarts (qt)
1 liter (l) = 0.26 gallon (gal)
1 cubic meter (m^3) = 36 cubic feet (cu ft, ft^3)
1 cubic meter (m^3) = 1.3 cubic yards (cu yd, yd^3)

TEMPERATURE (EXACT)

$[(9/5)y + 32]$ °C = x °F

QUICK INCH-CENTIMETER LENGTH CONVERSION

INCHES	0	1	2	3	4	5	6	7	8	9	10

CENTIMETERS 0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 | 25.40

QUICK FAHRENHEIT-CELSIUS TEMPERATURE CONVERSION

°F	-40°	-22°	-4°	14°	32°	50°	68°	86°	104°	122°	140°	158°	176°	194°	212°
°C	-40°	-30°	-20°	-10°	0°	10°	20°	30°	40°	50°	60°	70°	80°	90°	100°

For more exact and/or other conversion factors, see NBS Miscellaneous Publication 286, Units of Weights and Measures. Price $2.50. SD Catalog No. C13 10 286.

TABLE OF CONTENTS

TABLE OF CONTENTS (continued)

LIST OF ILLUSTRATIONS

LIST OF TABLES

LIST OF ILLUSTRATIONS

Figure		Page

LIST OF TABLES

Table | Page

ACRONYMS

ACARS	Airborne Communication and Reporting System
ADI	Attitude Director Indicator
AGL	Above ground level
AI	Artificial Intelligence
AP	Autopilot
AQP	Advanced Qualification Program
ATC	Air Traffic Control
ATIS	Automatic Terminal Information Service
BTA	Behavioral task analysis
C	Captain
CAT	Category
CDM	Critical Decision method
CL	Checklist
COA	Course of Action
CP	Co-Pilot
CRM	Crew Resource Management
CTA	Cognitive Task Analysis
EICAS	Engine Indication Crew Alerting System
EPR	Engine Pressure Ratio
FAA	Federal Aviation Administration
FAF	Final Approach Fix
FAR	Federal Aviation Regulation
FL	Flight Level
FMS	Flight Management Systems
FO	First Officer
GA	Go Around
GS	Glide Slope
ILS	Instrument landing system
ISD	Instructional Systems Design
KA	Klein Associates Inc.
KIAS	Knots Indicated Airspeed
LNAV	Lateral Navigation
LOC	Localizer
LOFT	Line Oriented Flight Training
MCP	Mode Control Panel
PF	Pilot flying
PMI	Performance measurement instrument
PNF	Pilot not flying
PPDR	Pilot Performance Description Report
SAL	Southeast Air Lines
SD	Standard Deviation
SE	Standard Error

SFAR	Special Federal Aviation Regulation
SID	Standard Instrument Departure
SME	Subject Matter Expert
SOP	Standard Operating Procedures
SPO	Supporting Proficiency Objective
TOGA	Takeoff - Go Around
TPO	Terminal Proficiency Objective
UNSAT	Unsatisfactory
VNAV	Vertical Navigation

1. INTRODUCTION

With the introduction of Special Federal Aviation Regulation (SFAR) Number 58, the Federal Aviation Administration (FAA) provided for approval of an alternative method for qualifying, training, and certifying individuals under FAR parts 121 or 135. The Advanced Qualification Program (AQP) allows air carriers to restructure the way that they train flight crews and the way that they develop training programs. Participation in AQP is voluntary, but carriers may realize several benefits from implementing an AQP: enhanced flight crew proficiency, cost savings through more effective and efficient training, and the ability to tailor training to the specific needs of the individual carrier.

Those groups working toward implementing an AQP face many considerable hurdles. First, AQP constitutes a substantial deviation from the way many carriers have developed and conducted training in the past. Carriers have invested billions of dollars on their existing training programs; they have confidence that these programs are effective. In some cases, current training programs and departments have been in existence for so long that they have become entrenched, extremely resistant to change. For example, years ago someone made a distinction between ground school and flight training. This distinction has evolved to the point where these two types of training are subsumed under different divisions of the company, they adhere to different training philosophies and compete for limited training resources. Often, the ground-training and flight-training divisions lose sight of the fact that they share common goals. Successful implementation of fully integrated training under AQP will require that these organizational barriers be lowered if not eliminated altogether.

A second problem internal to some carriers is that, in most cases, they possessed very little expertise in instructional design. Few carriers entered the program with extensive expertise on board.

In addition, there are no existing models for AQP development in the airline industry. That is, the carriers cannot simply build an AQP by copying one already in existence or by purchasing a turn-key training program. The military has used the Instructional Systems Design (ISD) for some years, but there are few large scale applications of ISD in the civilian aviation industry. Thus, each initial AQP applicant must begin from the ground and build up. Subsequent applicants may be able to build on the lessons learned by the initial group of carriers developing AQPs.

Several hurdles to AQP concern adapting the instructional design process to the task of flying airplanes. There are many applications of instructional design, but most of them involve relatively simple jobs. In addition, they often involve *ab initio* training programs where the student is completely inexperienced. Thus, much of the literature about instructional design focuses on training novices to perform relatively simple tasks. Neither of these conditions exists in commercial aviation. First, flying airplanes is not a simple task; it involves a complex mix of perceptual, motor, and cognitive skills. Second, very few commercial carriers in the U.S. conduct *ab initio* training. Even students enrolled in new-hire training programs are experienced aviators. AQP applicants must modify instructional design techniques to accommodate these domain specific conditions.

Finally, AQP applicants must address the dynamic nature of technology in the aviation industry. Recent years have witnessed dramatic changes in the business of flying. Training developers must ensure that they can adapt training objectives and performance standards to these evolving technologies that include hardware and software as well as advances in our knowledge of team processes and performance.

Clearly, AQP will require a significant investment of resources from the carriers. The initial curriculum development should prove the most expensive as the program developers discover the most effective process before they realize any benefits from the AQP. However, both the FAA and the carriers should learn at each step of the process and subsequent development efforts should be more efficient. The FAA does not place the entire burden of AQP development on the carriers. In fact, the FAA views AQP as a cooperative process between the government and industry and is willing to support the carriers' developmental efforts.

The present research is a product of the FAA's intent to provide such support. The FAA contracted Klein Associates Inc. (KA) to conduct research that would lead to the development of reliable and valid instruments for measuring crew performance. The purpose of this report is to describe KA's research under this contract.

This report presents the research findings in several sections. The Introduction discusses background material for several areas that are relevant to this work including the AQP process, the integration of crew resource management (CRM) and technical skills, and the elicitation of knowledge from experts. The Methods section describes how we conducted the work. The Results section describes the findings of the study. Finally, the Discussion section presents some conclusions and observations that we have drawn from conducting this work.

1.1 Advanced Qualification Program

AQP is an alternate method of qualifying, training, and certifying flight crew members and others subject to training and evaluation requirements of Federal Aviation Regulation (FAR) parts 121 and 135. Special FAR (SFAR) 58 (FAA, 1991a) provides for the approval of an AQP. Advisory Circular 120-54 (FAA, 1991b) describes an acceptable method for developing, implementing, and maintaining an AQP.

Training and qualification requirements for flight crew members have not changed significantly since 1970. Most aviator training programs remain time and event based. That is, students progress through the programs by completing specific training events and by receiving prescribed amounts of training time in various devices. For example, a curriculum may require a student aviator to complete 20 hours of training in a flight simulator. In some cases, the curriculum may specify the events that the student is to accomplish during those 20 hours, but not the training goals. In other cases, the curriculum may not prescribe what the student is to accomplish during those 20 hours or the level of proficiency that s/he should attain through such training.

Since 1970, we have seen significant advances in technologies that affect what aviators must learn and how they can learn. These include both aircraft and training technologies. Aircraft automation dramatically impacts flying tasks and alters the roles of the flight crew. Numerous researchers have documented the effects of automation on the flight crew (e.g., Hutchins, 1991; Norman, 1991; Wiener, Chidester, Kanki, Palmer, Curry, & Gregorich, 1991). Automation improves the ability of the flight crews to perform traditional flight tasks such as instrument approaches. But, in many cases, automation alters the roles of the crew members and the cognitive processes required to accomplish tasks successfully. In fact, automating functions leads to more complex cognitive tasks (Howell & Cook, 1989). Previous methods of training and training development have difficulty accommodating such changes.

In addition, we have seen significant advances in training technology. For example, the advent of flight-training devices and flight simulators have dramatically altered the way that we prepare student aviators for flight operations. Computer-based training, interactive technology, and the advent of systems approaches to training development have had similar impacts on our ability to conduct effective training efficiently.

The intent of AQP is to allow training departments to develop innovative training and qualification programs that take advantage of developments in training technology, methods, and techniques that will enhance professional skills beyond the level required by current standards. The cornerstone of AQP is true proficiency-based training and qualification. Student pilots progress through the training programs based on demonstrated proficiency at critical milestones. Progression does not depend solely on accomplishing specified events or numbers of training hours. SFAR 58 describes five phases of an AQP:

- Initial application
- General curriculum development
- Training system implementation and courseware development and implementation
- Initial operations
- Continuing operations

The remainder of this section describes each of the five phases of the AQP process.

In the initial phase, the applicant establishes their intent and approach for developing an AQP. The applicant submits a complete plan for approval by the FAA. The plan must fully describe how the applicant will develop, implement, and manage the proposed AQP. FAA review ensures that the applicant's plan conforms to guidelines and will accomplish the desired goals.

After approval of the initial application, the carrier enters Phase II. Phase II comprises three major steps: the development of proficiency objectives and qualification standards, development of a complete training curriculum and syllabus, and development of the training resource requirements and a plan for implementation and operation. The FAA must review and approve the carrier's program at each step.

The primary goal of Step 1 is to develop Terminal Proficiency Objectives (TPOs) and Supporting Proficiency Objectives (SPOs) that will provide a basis for curriculum development. Proficiency objectives define each task that students must learn and provide a basis for evaluating student performance. A complete set of proficiency objectives fully describes the job performed in a particular crew duty station. The proficiency objectives must include performance statements, equipment and environmental conditions that affect performance, and standards that define satisfactory performance. The applicant develops these objectives from a thorough analysis; the FAA recommends that the applicant conduct task and subtask analyses to accomplish this goal. Tasks are organized into TPOs, while subtasks are organized into SPOs.

Step 2 of Phase II comprises the development of a course curriculum based on the results of Step 1. A curriculum is the learning order sequence of segments, modules, lessons, and lesson elements and provides a means of planning hours, media, methods, and scenarios to be used. Developing a curriculum includes extracting procedural and cognitive skills, knowledge, and attitudes necessary to accomplish each SPO and allocating these to one or more curriculum segments. Next, the applicant establishes a learning hierarchy for each segment's objectives and skills, develops sequenced lessons for each segment, and groups the lessons into basic modules by subject or purpose.

In Step 3 of Phase II, the applicant determines the resource requirements for the AQP. The applicant develops two documents, one that details training resource requirements and another that describes the AQP implementation and operation plan. As part of the training resources, the applicant describes the training facilities, the courseware, the number, type, and qualification of instructors, the evaluators, and the plans for ensuring the quality of the program data and the performance measurement data.

The plan for implementation and operation includes both formative and summative evaluation plans. The formative plan describes evaluations of all of the elements identified in the requirements document including small group tryouts of all new courseware, software, and equipment. The summative plan describes the plan for evaluating the AQP during Phase IV.

In Phase III, the applicant completes the final steps prior to initial implementation of the AQP. It includes four steps:

- Develop and implement courseware and testing materials
- Implement the FAA-approved Formative Evaluation Plan
- Train, evaluate, and qualify instructors and evaluators

- Review and modify the Summative Evaluation Plan and the AQP Maintenance Plan using the Formative Evaluation Plan

Following successful completion of these four steps, the FAA provides initial approval of the AQP and the applicant may begin initial AQP operations.

The operator initiates these operations as Phase IV. Under Phase IV, the operator implements the entire AQP training and evaluation cycle and the AQP Maintenance Plan. In addition, the operator continues the summative evaluation to ensure the effectiveness and appropriateness of all aspects of the program.

In Phase V, the operator continues utilization of the AQP. The principal task of this phase is to monitor the proficiency of the students, instructors, and evaluators as a continuing assessment of the program's effectiveness. In addition, the operator modifies and assesses the program as required by changes in equipment or students.

AC 120-54 (FAA, 1992) describes a complete AQP, one that incorporates three curricula for each duty position in each make, model, and series (or variant) aircraft. The curricula are indoctrination, initial qualification, and continuing qualification. Developing these programs and putting them in place represents a considerable amount of effort and presents numerous pitfalls. Therefore, the AQP applicants have chosen to begin by working on a single piece of the AQP puzzle. For example, one applicant is developing a single curriculum, continuing qualification, for a single aircraft. After completing this curriculum, this group will build on their successes by adding other curricula and aircraft. This AQP will not be approved, however, until all curricula are complete.

The heart of AQP is true proficiency-based training and evaluation. Student pilots progress through training based on their ability to achieve specified objectives at each step. Thus, it is critical that AQP applicants develop reliable and valid instruments to measure pilot and crew proficiency relative to the training objectives. Trainers must be able to accurately determine when a pilot's skills meet or fail to meet the training objectives. Accurate performance measurement is important for two reasons. First, student progression through the curriculum depends on his/her ability to meet the objectives. Second, the accumulated documentation of student performance serves as the primary measure of the AQP's effectiveness. That is, performance measurement instruments serve as a means for evaluating the training program as well as the individual pilot or crew.

1.2 Integration of CRM and Technical Skills

Crew Resource Management grew out of the recognition that the majority of aviation accidents and incidents were attributed to the ubiquitous root cause "human error." However, closer inspection revealed that the principal contributor to human error accidents was the failure of crew members to utilize all of the resources available to them and to perform effectively as a team (see Cooper, White, & Lauber, 1979). Since that time, the aviation community has increasingly accepted CRM concepts. Many researchers, particularly those at NASA Ames Research Center and the University of Texas, have worked to refine our knowledge of CRM and to develop training principles designed to improve the effectiveness of teams comprised of individual pilots. A variety of CRM training courses have evolved from these research findings; some version of CRM training can be found in virtually all segments of flight training.

A thorough discussion of CRM is beyond the scope of this report. However, a few points warrant discussion here. The following paragraphs present a few relevant points followed by a discussion of efforts to integrate the training and measurement of CRM skills with those of technical flying skills. There is a considerable literature about CRM and CRM training; the following is a short list of primary references:

- Cooper, White, and Lauber (1979)
- Foushee & Helmreich (1988)
- Helmreich (1984)
- Helmreich (1987)
- Helmreich, Foushee, Benson, and Russini (1986)
- Helmreich, Wilhelm, Gregorich, & Chidester, (1990)
- Povenmire (1989)
- Ruffell Smith (1979)
- Sams (1987)

1.2.1 Crew Resource Management Research

Recently, the FAA adopted the term Crew Resource Management to replace Cockpit Resource Management (see FAA, 1992). This change reflects expansion of the concept of crew to include cabin crew, dispatch, air traffic controllers, and maintenance personnel in addition to the flight crew. In this report, CRM refers to the expanded concept.

Foushee and Helmreich (1988) described group performance in terms of three sets of factors: input, process, and output. Input factors include the personal characteristics that individuals bring to the group, characteristics of the group itself, and environmental and task factors. Process factors include the dynamics of group interaction, the mechanics of group functioning, and how group members communicate with each other. Outcome factors refer to those concerning how well the group performed the task and interpersonal consequences (e.g., changes in member satisfaction, attitudes, and cohesiveness).

The main body of CRM research has explored input factors and their effects on crew performance. Other input factors include the structure of the training task and the role structure in the cockpit. However, these are not relevant for the present study and will not be discussed further. The primary focus has been on how personality characteristics and attitudes affect the group process. Numerous researchers have demonstrated the effects of personality traits on group performance (e.g., Fiedler, 1967; Helmreich, 1982). Helmreich showed that traits clustering in the areas of achievement motivation and interpersonal achievement have strong associations with measures of group performance. Helmreich (1986) later argued that personality traits are enduring characteristics that are resistant to change. Chidester, Helmreich, Gregorich, and Geis (1991) described personality as a barrier limiting the effectiveness of CRM training. As such, personality traits in the pilot population are not amenable to manipulation by training. Rather, they can be manipulated by selecting for specific traits during the initial hiring process.

Wheale (1983) and Foushee (1984) reported that attitudes about flight deck management affect the quality of resource management and subsequent crew performance. Since then, other studies have verified these links and identified the relevant attitudes. For example, using the Cockpit Management Attitudes Questionnaire, Helmreich, Wilhelm, and Gregorich (1988) identified attitude profiles that distinguish pilots rated as superior from those pilots rated as below average. In addition, Helmreich (1984) has described attitudes as malleable and subject to change through training. He further reported that changes in CRM practices will occur only if the attitudes of crew members toward crew interactive behaviors change. Thus, pilot attitudes about resource management remain a primary target for change through training.

In contrast to the input factors, process factors have received little attention. Most of the research on process factors has investigated the amount and patterns of verbal communications between crew members. Very little is known about other group processes. Furthermore, this work has a somewhat different flavor. Early CRM studies established the link between good resource management and enhanced aviation safety. However, system safety is a cumbersome criterion measure to use, particularly for short duration studies. The effects of any manipulations on system safety require significant periods of time to manifest themselves. Rather than investigate the effects of CRM manipulations on safety, some researchers began studying relationships between communication processes and outcomes such as technical performance and mission accomplishment. The principal assumption in these studies is that good resource management should be manifested in successful outcomes. That is, crews that manage their resources better should be more successful and certain patterns of communication (a group process) should be related to crew success.

The studies of verbal communications among crew members have investigated the relationships between communication patterns and crew performance variables. Several studies have demonstrated a link between patterns of verbal communication and crew performance on transport flying (e.g., Foushee & Manos, 1981; Kanki, Lozito, & Foushee, 1987; Orasanu, 1990). Others have identified patterns of communication that predict performance on measures of effectiveness for military missions (e.g., Krumm & Farina, 1962; Povenmire, Rockway, Bunecke, & Patton, 1989; Thornton, Kaempf, Zeller, & McAnulty, 1991). These studies have produced recommendations about types of communication that support good performance and should be encouraged between crew members. For example, members of successful crews tend to acknowledge receipt of information or directives from the other crew member. Thus, CRM training programs encourage pilots to verbally acknowledge when they receive information and directives.

1.2.2 Crew Resource Management Training

Developers of CRM training have adhered to the notion that manipulating input factors is the most effective means of affecting crew performance on outcome factors (Foushee & Helmreich, 1988). That is, adjusting pilots' attitudes in the training environment will improve subsequent performance in the operational environment. Training programs focus on establishing appropriate group norms through training, while modifying the pilot population through the initial selection process.

The identification and training of crew resource management skills have progressed considerably since the late 1970's. Over the last decade, researchers have identified reliable behavioral markers of CRM, and training developers have created numerous training programs. Most flight-training centers (military and civilian) now incorporate CRM modules into their curricula. Recognizing the importance of CRM training for safe flight operations, SFAR 58 (FAA, 1991a) prescribes that all AQPs will provide CRM training in both qualification and continuing qualification curricula. Several reviews describe various CRM training strategies and training programs that have been implemented (e.g., Povenmire, 1989; Sams, 1987).

Typical CRM training modules include both classroom and simulator training. Classroom CRM instruction generally concentrates on the dimensions thought to influence crew performance such as interpersonal skills, leadership-followership, communication, and stress management. Training programs may include classroom presentations, practice and feedback, and reinforcement of learning (Foushee, 1985). These programs typically attempt to create a collegial atmosphere in which crew members share problem-solving and decision-making duties. However, developing a collegial atmosphere does not train individuals or teams to make decisions. Most CRM training focuses on these environmental types of skills without addressing the process skills.

In addition to classroom instruction, many training centers now use line oriented flight training (LOFT) for crews to enhance their CRM skills. Crews plan and fly operationally realistic missions in a flight simulator. The LOFT scenario usually includes equipment malfunctions or unusual conditions not normally seen in line flying. The flights occur in real time with no interventions. They are recorded for debrief immediately following the flight. During the debrief, the crew reviews the recordings and conducts a self-critique with the help of the instructor/facilitator.

Typically, crews are not evaluated during LOFT missions. Facilitators attempt to create an open environment conducive to learning and exchanging ideas. The objective is to provide a rewarding experience for the crews and to create an environment in which pilots will openly seek to improve their skills. The threat of evaluation or the perception of any risk to the pilots would reduce the effectiveness of the training sessions.

CRM instructors/facilitators teach pilots about several dimensions of CRM behavior, areas in which they can improve their resource management. These dimensions vary from one training program to the next, but they all contain similar items. The following is a list of dimensions extracted from one air carrier's CRM training program:

- Briefing
- Inquiry, advocacy, and assertion
- Crew self-critique
- Communications and decisions
- Leadership-followership and concern for tasks
- Interpersonal relationships and group climate
- Preparation, planning, and vigilance
- Workload distribution and avoiding distractions.

While observing crews perform their LOFT mission and during the debrief, the instructor/facilitator concentrates on these dimensions. Specific behavioral markers reflect performance on each of these dimensions. Appendix A presents all of the markers for each of the eight dimensions listed above (Helmreich, Wilhelm, Kello, & Taggant, 1987). Facilitators look for these markers during the LOFT mission and use them to anchor the self-critique during the mission debrief session.

1.2.3 Development of CRM Integration

Two types of flying skills have evolved, technical flying skills and CRM skills. To date, they have been kept separate. Technical skills have evolved over the decades of flying. Aviators have developed both explicit and implicit skills and knowledge. These have served as the bread and butter of flight-training programs. In contrast, knowledge about CRM and how to train CRM has developed over the last 12 years and continues to evolve. The importance of these skills and their impact on safety and performance has become clear. However, training and evaluating the two types of skills have remained segregated to a large degree.

Many aviators intuitively understand the positive impact that good CRM skills have on technical performance. In fact, many experienced aviators claim that CRM is nothing new to the cockpit--that good aviators have been doing it for years. The aviation community recognizes the need for proficiency on both sets of skills. Two pilots may have expert stick and rudder skills, but not be able to work together as a team in the cockpit. This inability to work as a team degrades their ability to fly the airplane; such mistakes are manifested as a variety of technical errors. Crews violate altitude or airspace restrictions, they run out of fuel, they have near misses, they crash airplanes. That is, CRM errors by themselves are not critical. However, the technical errors that they create can be catastrophic. Postmortem analyses of the event sequences leading to aviation accidents and incidents invariably reveal the contributions that CRM errors make to technical errors.

Despite the links between CRM and technical skills, researchers and trainers have treated the two separately. Training programs isolate CRM into separate modules in which pilots learn about the dimensions of CRM and participate in exercises designed to modify their attitudes. Even LOFT sessions separate the two sets of skills. Crews perform LOFT missions in the context of operational flying, but their attention is on CRM. Instructors ensure that the crews recognize and evaluate their own CRM skills; the instructors do not train technical skills. Typically, sufficient time is not provided to review and practice technical skills.

It might be more profitable to argue that we should not delineate between CRM and technical skills. The ultimate manifestation of CRM skills is in technical performance. Poor CRM may be reflected in technical errors. Good CRM contributes to uneventful, safe flying. For example, the poor CRM crew fails to manage their time and workload adequately. Therefore, they leave too much work to be done as the crew is on the approach. They get rushed because there is much to do and very little time. Because they are rushed, they truncate or miss checklists, miss radio calls, or maybe forget to cross check the settings of altimeters and speed bugs. This is under normal operations. Introduce an abnormal condition, and the crew is unable to bear the additional work in the same period of time.

In addition, even though it is accepted that good CRM skills are vital for successful and safe crew performance, CRM skills are not considered during evaluations of pilot proficiency. The FAA does not require that check airmen assess pilots on their CRM proficiency. This is probably due to two factors. One is a tradition created in CRM training. Early trainers believed that pilots would accept CRM concepts more quickly and that CRM training courses would be more effective if conducted without the threat of evaluation. The course would effect positive attitude changes if the pilots did not have to put their certifications on the line. Second, CRM had not matured to the point where we could reliably measure CRM performance. Performance measurement technology lagged behind the development of CRM concepts. Many argue that we cannot objectively assess performance on the "soft skills." Both of these factors have combined to create a lasting impression among many aviators that they should not be evaluated on CRM proficiency. Helmreich, Hackman, and Foushee (1988) discuss the problems with crew performance evaluations.

These beliefs may have been reasonable reactions to the first generation of CRM training. However, our ability to construct valid and reliable measures of CRM performance and to integrate training of CRM skills with technical skills is changing. It improves as we learn more about CRM concepts and the mechanisms of team performance. Technology has advanced to the point where we must now begin to tackle the problems of integration and measurement. The FAA motivated progress in these areas by suggesting that AQPs integrate CRM training into technical training using proficiency objectives and objective measures of CRM proficiency (FAA, 1991). In response, the Air Transport Association developed a focus group for CRM integration that serves as part of a subcommittee for AQP. In addition, the FAA sponsored a workshop among industry, research, and government parties to assess the state of CRM integration and to identify future directions (Lofaro, 1992). Several researchers have begun developing techniques for integrating CRM and technical skills. Smith's (1991) Crew Performance Model is the most prominent attempt in the air carrier industry. The following paragraphs describe this model.

Smith (1991) developed an analytical method for identifying the CRM and technical skills required to perform specific tasks under specific condition sets. A hierarchical model of crew performance underlies the method. Figure 1-1 illustrates the model hierarchy. The model describes successful crew performance as a product of two types of factors: technical skills and human factors skills. Each set of factors is further divided into four clusters, clusters into categories, and categories into behavioral markers. The technical performance clusters comprise the skills needed for basic operation of the aircraft including Flight Maneuvers and Attitude Control; Propulsion, Lift, and Drag Control; Systems Operations; and Malfunction Warning and Reconfiguration. Smith derived markers for technical performance from a variety of aircraft operations manuals, company procedures, and FAA documents.

The human factors skills are of particular importance here. Three of the four clusters incorporate CRM skills: Communication Process and Decision Behavior, Team Building and Maintenance, and Workload Management and Situational Awareness. These clusters are represented in the FAA's CRM advisory circular (FAA, 1992). The categories of skills grouped by cluster correspond to the eight dimensions of CRM described above. The behavioral markers for these skills are those presented in Appendix A. The number of markers grouped under any one category varies between five and ten.

Smith proposed that this model serve as the framework for developing training objectives that integrate CRM with technical performance. To implement the model, a subject matter expert (SME) begins by selecting a specific task and set of environmental and equipment conditions. The SME then constructs a matrix bounded by the eight clusters and the subtasks of the particular task. Subtasks are derived from a behavioral task analysis conducted previously. The SME uses his/her experience as a pilot to identify which clusters are relevant for each subtask. That is, the SME walks through the task and notes each of the behavioral requirements (technical and human factors) for the given condition set. The SME uses this analysis to produce an exhaustive list of the behavioral markers that apply to each subtask for the given condition set. Thus, the list integrates markers of technical and CRM performance.

Smith's approach has two major limitations. First, the behavioral markers are too general for developing specific performance statements and performance standards. Second, Smith does not identify which methods should be used to elicit knowledge from SMEs. Both of these are discussed below.

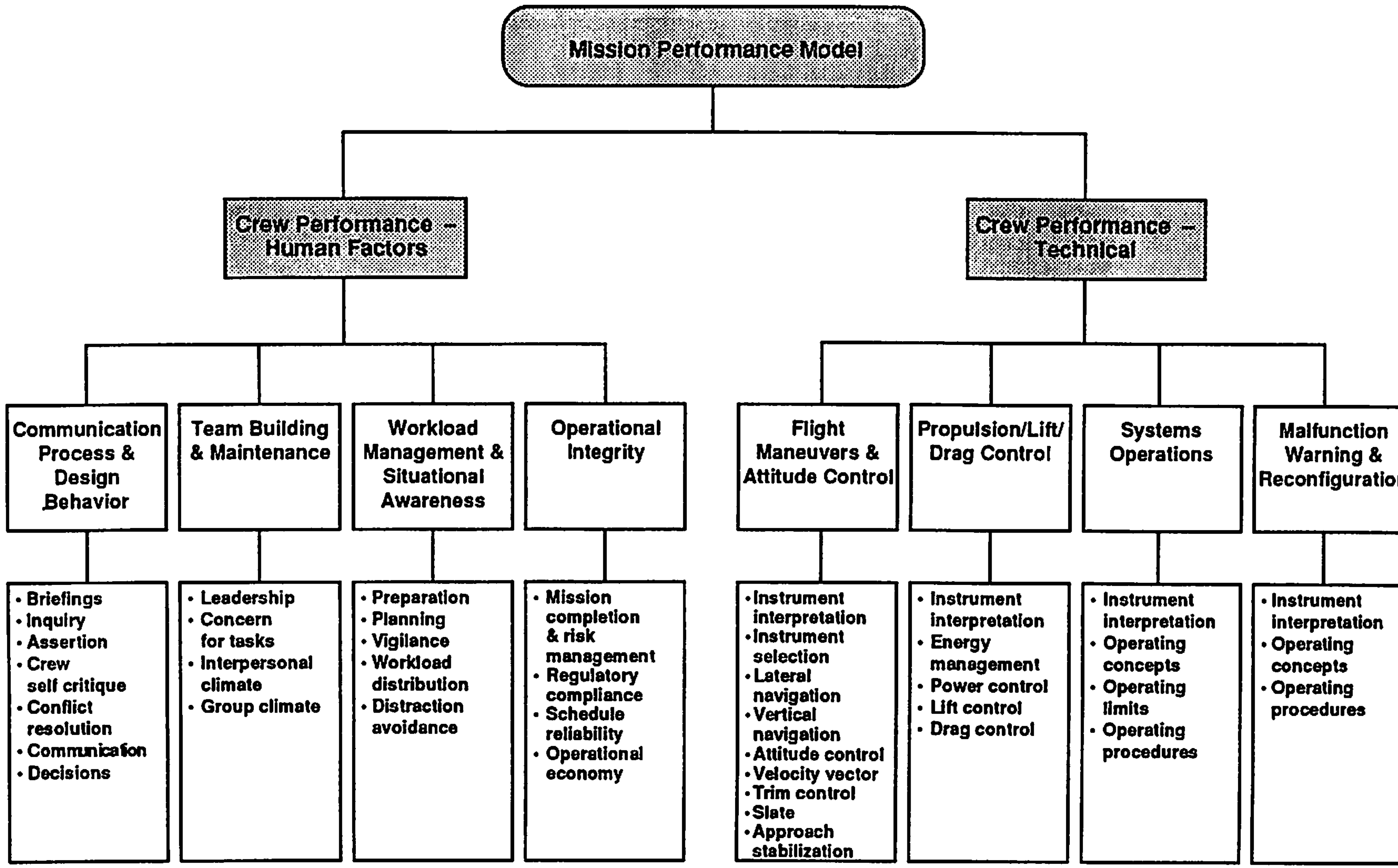


Figure 1-1. Model of Crew Performance (Smith, 1991)

The behavioral markers are too general for developing adequate performance statements and performance standards. The behavioral markers are not specific behaviors, rather they represent classes of behaviors. For example, "Coordinates flight deck activities" subsumes a wide variety of specific behaviors that vary depending on the task at hand and the condition set. The CRM facilitator looks for the presence of many specific behaviors that indicate the crew is coordinating its flight deck activities. It is this next lower level of specific behavior that is essential for comprehensive performance statements.

Smith does not describe how to elicit information from the SME analyst. Researchers have reported success with a variety of knowledge elicitation techniques designed for use with domain experts. These techniques include various scaling methods, protocol analysis, job samples, observations, and interviews. Eliciting the appropriate information from experts becomes a formidable problem as the nature of flying evolves from individual technical skills to team and cognitive skills. As aircraft become more complex, overt behavioral actions of the pilot remain important, but the covert cognitive tasks of the pilot become increasingly complex and important. Gaining access to these cognitive tasks requires that the analyst use a different set of tools. This problem is of direct relevance to this project and is discussed in more detail in the following section.

1.3 Knowledge Elicitation

Techniques for knowledge elicitation have received considerable attention in the areas of instructional design and artificial intelligence (AI). Both areas have the same objective: to codify what the expert knows. Instructional designers need models of expert knowledge to build a pathway that will guide trainees from novice performance to expert. AI researchers need models of expert knowledge to guide development of "smart systems." They attempt to build systems that contain knowledge similar to that held by domain experts.

Behavioral task analysis (BTA) is the basic building block of instructional system design. Designers base every step of the process on the results of these analyses. Furthermore, BTAs rely on the ability of designers to extract specific knowledge from domain experts. BTAs provide a very detailed description of how an expert performs the task in question by focusing on the behavioral components of the task. A variety of techniques that focus on overt behaviors are appropriate for eliciting this type of knowledge from domain experts including observations, interviews, and protocol analyses. Designers query domain experts to create a hierarchical listing of the physical requirements of the task. Cognitive or perceptual-motor requirements (e.g., decide, assess, judge) may enter the analysis, but usually at the lowest level of the hierarchy. Thus, BTAs are an accurate description of the physical task, but generally fail to adequately capture complex cognitive requirements of the task.

BTA is adequate for deriving training for tasks with fixed procedural sequences (e.g., assembly line tasks) and largely psychomotor skills. However, it has proved inadequate for developing training for tasks that have large decision-making components and complex cognitive requirements (see Redding, 1989; Ryder, Redding, & Beckshi, 1987). Other techniques are required to build adequate models of cognitive performance.

Researchers suggest that Cognitive Task Analyses (CTA) be integrated with BTAs to provide a complete picture of expert performance (see Redding, 1989, 1990; Ryder et al, 1987; Schlager, Means, & Roth, 1990). CTAs have been used effectively in both instructional design (Redding & Lierman, 1990; Schlager, Means, & Roth, 1990; Schneider, Vidulich, & Yeh, 1982) and for systems design (e.g., Kaempf, Wolf, Thordsen, & Klein, 1992; Roth & Woods, 1990; Woods & Hollnagel, 1987). The goal of CTA for instructional design is to delineate the mental processes and skills needed to perform a task at high levels of proficiency. In addition, CTAs frequently model the changes in cognitive structures that occur as a trainee advances from novice to expert performance. These processes include the conceptual and procedural knowledge and mental models used to support task performance.

CTAs have the same reliance on domain experts as BTAs; knowledge held by experts is the basic component of both types of analysis. However, the different objectives of the two types of analyses demand different methods for data collection. Various researchers have reported success with focused interviews and psychological scaling procedures such as sorting, recalling or rating task-relevant knowledge, measuring inter-response time, response times, output order, errors, and similarity judgments.

Regardless of the type of analysis used, the literature points to the difficulty eliciting certain types of knowledge from experts and the nature of the knowledge held by experts (Chi, Glaser, & Farr, 1988; Hoffman, 1987). However, these and other researchers have had considerable success with techniques specifically designed to elicit these types of knowledge (e.g., Klein, Calderwood, & MacGregor, 1989). Experts often have extreme difficulty verbalizing much of what they know. That is, they hold their expert knowledge implicitly. This is partially because they have lost contact with the verbal forms of the knowledge and partially because they never held specific pieces of knowledge in a verbal form.

Part of the nature of expertise is that the expert performs many tasks automatically. They once learned how to perform a task perhaps through books or lectures. However, after repeated practice performance becomes routinized and they no longer remember the individual steps or how to perform the task. Thus, they once knew a verbal description of the task, but have since forgotten it. An interviewer needs to design specific probes to bring this type of knowledge back to the front. For a thorough discussion, see the learning theories of Anderson (1982) and Fitts and Posner (1967). Experts also possess some knowledge which never began in a declarative form (Berry, 1987). This type of knowledge arises from an implicit learning process, particularly for perceptual learning. Since this knowledge never had a verbal component, experts have difficulty finding the words to describe it. Likewise, the interviewer must have well-designed probes to help the expert articulate this type of knowledge.

When learning to fly a helicopter close to the ground, student pilots must develop an awareness of numerous sensory cues. For example, they must learn to judge rate of closure by visually observing vertical and horizontal movement. Instructors can tell students what the important dimensions are, but they cannot transfer to students knowledge of what the accurate sight picture looks like. Only through repeated practice can students learn to "see" the correct picture. Since they did not learn the knowledge of sight picture through words, they have difficulty describing correct sight picture when asked.

Much of our early language acquisition provides another example of the implicit learning process. Toddlers pick up many rules of grammar by observing and listening to others. Adults do not tell children to form past tense verbs by adding "ed" to the end. Yet, English-speaking children easily acquire this knowledge. In fact, implicit procedural knowledge held by toddlers is evident when they try to generalize rules by forming past tense versions of irregular verbs. A toddler might use the word "comed" instead of "came." Presumably, children learn these rules by observing and listening to others in their environment.

The point of this discussion is that experts cannot be expected to spew forth information that differentiates their performance from that of novices without some well-conceived approach to knowledge elicitation. Frequently, experts do not know what is needed by the analyst nor can they verbalize much of the knowledge that makes them distinctive. This is particularly true for cognitive and perceptual knowledge. A successful analysis depends on the ability of the analyst to construct an effective knowledge elicitation procedure tailored to fit the specific domain and the goals of the project.

The AQP process appears to lack these well-focused knowledge elicitation procedures. Most applicants are using standard BTA techniques that focus only on overt behaviors. These techniques are inadequate for identifying either team performance or the cognitive components of performance.

1.4 This Project

The preceding discussion highlighted some of the issues critical to the AQP process. The research described in this report was designed to address these issues.

The project had two specific objectives. The first was to determine the utility of the Critical Decision method (CDM) as a knowledge elicitation tool for developing measures of performance. The second and more important goal was to develop and evaluate a prototype performance measurement instrument that integrates the measurement of CRM proficiency with that of technical proficiency.

An overarching goal of this project was for the research to provide immediate support to AQP applicants during their development process. Thus, it was important for the researchers to work directly with one of the applicants. The FAA coordinated a working relationship between the researchers and a major air carrier. We conducted this research within the context of one applicant's AQP. The applicant is a major U.S. carrier that we will call Southeast Air Lines (SAL) for the purposes of this report. SAL provided technical support including subject matter experts (SMEs), flight crews, flight simulators, and extensive cooperation from their AQP development team. The FAA provided funding and technical guidance.

SAL's AQP development efforts focus on the continuing qualification curriculum for the B-757/767 aircraft. These two aircraft are operated by pilots with a common type rating. Thus, they are subsumed under the same training programs. After completing this initial AQP curriculum, SAL will expand AQP development to other curricula and aircraft.

The nature and status of SAL's AQP development had several effects on this project. First, pilots enrolled in recurrent training provide a different challenge than *ab initio* or initial-qualification students. Pilots in recurrent training are fully qualified and have extensive experience. They maintain proficiency on most skills through the course of operational flying. Thus, recurrent training is designed to enhance and assess their proficiency on a selected set of skills that may degrade during the recurrent interval.

Second, SAL uses several LOFT scenarios for recurrent training. Every pilot enrolled in recurrent training must complete a LOFT session. Thus, we could incorporate an evaluation of the prototype performance measurement instrument into ongoing recurrent training. This conserved resources in two ways. We did not have to request additional flight crews and instructors to serve as subjects in the evaluation, and we did not have to expend resources to design a LOFT scenario specifically for this project. Therefore, we identified a LOFT scenario currently used by SAL, selected target tasks incorporated in that scenario, and conducted the entire study with the ultimate goal of evaluating the PMI in the context of this LOFT scenario.

Finally, developing measures of proficiency requires task descriptions (usually a task analysis), performance statements, and performance standards. When we began this project, SAL had not completed these documents. We first turned our attention toward completing the front-end analyses that enabled us to develop prototype measures of performance.

Thus, this project served as an AQP development effort in miniature. For a select group of tasks, we analyzed the task requirements; we identified performance standards; we developed proficiency objectives; and, we developed and evaluated a PMI. For a select group of flight tasks and conditions, we:

- Identified the relevant tasks and subtasks
- Identified performance standards
- Developed terminal and supporting proficiency objectives
- Developed a prototype performance measurement instrument
- Evaluated the effectiveness of the PMI

We performed these tasks to build straw men examples. We do not expect the task requirements or the proficiency objectives to survive long as AQP development efforts progress. Better ways to perform these tasks will be found. However, this process allowed SAL to examine the issues and pitfalls of each step prior to attempting them all on a larger scale, and enhanced their own analytical products.

2. METHODS

This section describes how we conducted the research. We progressed through the project in several steps including considerable initial effort to learn about SAL's operations and the B-757/767 aircraft. The following subsections describe each of the project's major steps including:

- Selecting flight tasks
- Eliciting knowledge from domain experts
- Pretesting the performance measurement instrument
- Evaluating the performance measurement instrument

2.1 Flight Tasks

Line pilots must attend recurrent training every 12 months. SAL's recurrent training program for the B-757/767 aircraft comprises four training days spent at a training center. Typically, two pilots who require training are assigned to attend the course as a flight crew. Most often, they have never worked together. The first two training days include ground school instruction. On the third training day, the flight crew receives training on specific flight maneuvers and conducts a Line Oriented Flight Training (LOFT) mission. On the final recurrent training day, the flight crew receives a proficiency check ride including an oral examination.

SAL developed several LOFT scenarios for its B-757/767 recurrent training curriculum. Each of these scenarios contains two flight legs and simulates operational problems experienced by line flight crews. We chose to study tasks selected from these LOFT scenarios because it would enable us to incorporate evaluation of the prototype PMI into ongoing recurrent training. In addition, by using an existing scenario we avoided the expense of scenario development.

We selected tasks from the most recently developed scenario. With repeated exposures to the same scenario, pilots often learn what to expect next. However, very few instructors and none of the line pilots had flown this particular LOFT scenario. Therefore, the pilots participating in the evaluations would not know the order of events in the LOFT and could not anticipate upcoming events.

In the first leg of the scenario, the crew takes off from Cincinnati in poor weather and experiences an engine failure upon reaching 200 feet above ground level (AGL). The crew may choose to divert to another airport, but the most expeditious choice is to return to Cincinnati. The crew then performs a Category II ILS approach to Cincinnati with one engine functioning. Prior to reaching the decision height on the approach, Air Traffic Control (ATC) closes the runway and the crew must execute a single-engine missed approach. The crew then must determine an appropriate alternate airport and divert to that airport.

Most of the subject matter experts participating in this study felt that this scenario is a particularly difficult one. First, the engine failure occurs at a time that makes it technically challenging. At 200 feet AGL, aircraft yaw is significant because of the amount of power applied, and due to weather the crew has no outside visual references to maintain heading. Second, the scenario requires that the crew execute several procedures that they have not practiced in 12 months. Finally, the scenario forces the crew to deal with abnormal conditions and high workload early in the first leg. The crew does not have sufficient time to adjust to each other's work habits and personalities. Thus, the scenario puts the crew's team skills to the test very quickly.

This flight leg provides sufficient challenge for developing a performance measurement instrument. The leg includes difficult technical tasks as well as conditions and tasks that challenge the abilities of the crews to perform as teams. The leg includes the following tasks derived from SAL's task analysis:

- Takeoff with engine failure after V1
- Category II ILS Approach with single-engine operation
- Single-engine missed approach
- Divert to an alternate airport with single-engine operation.

2.2 Knowledge Elicitation

Interviews with subject matter experts provided data for the development of training objectives and a performance measurement instrument. We had two general objectives for the interviews. The first was to elicit information about the specific tasks and flight maneuvers that the flight crew must accomplish. The second was to elicit information about how expert instructors evaluate pilot performance that would serve as a basis for developing performance standards. The following two subsections describe our interview techniques and the interview procedures for this study.

2.2.1 Critical Decision Method (CDM)

CDM served as our principal means for eliciting knowledge from subject matter experts. CDM is a semi-structured interview technique derived from Flanagan's (1954) critical incident technique. Researchers at Klein Associates developed CDM under the sponsorship of the Army Research Institute for the Behavioral and Social Sciences to elicit the decision strategies used by experienced fireground commanders and emergency rescue personnel at the scene of a fire or emergency. Many of these decisions were found to depend on subtle perceptual cues and assessments of rapidly changing events that could not be articulated easily. Thus, probes had to be developed that would allow experts to focus on and describe aspects of their tasks that are normally only tacitly understood. CDM has been demonstrated to yield information richer in variety, specificity, and quantity than experts' typical verbal reports (Crandall, 1989). The method has been used in numerous studies and in domains as diverse as anti-air warfare, air battle management, fireground command, battle planning, critical care nursing, corporate information management, and fixed-wing and rotary-wing piloting (Brezovic, Klein, & Thordsen, 1987; Calderwood, Crandall, & Klein, 1987; Crandall & Calderwood, 1989; Crandall & Klein, 1988; Kaempf, Wolf, Thordsen, & Klein, 1992; Thordsen & Calderwood, 1989; Thordsen, Klein, & Wolf, 1990).

The CDM focuses on previous experiences of domain experts and applies a set of cognitive probes to elicit the experts' decision strategies, perceptual discriminations, pattern recognitions, expectancies, environmental cues, and errors. Initially, the interviewee relates a previous experience that proved tough or challenged his/her skills. In some ways, the CDM is a storytelling technique (see Deutsch, 1990; Schank, 1990) that is guided by the researcher. Often, experts are unaware of what they know and what they perceive; selected probes allow the researcher to elicit information about these factors within the context of the specific incident. Table 2-1 provides a list and explanations for the types of probes used in CDM.

The CDM is a semi-structured technique in that we do not ask prearranged questions in a prearranged order. Knowing the types of information that we want to elicit, the interviewer guides the expert through the incident. Each probe does not represent a single question, but a topic that the researcher will probe. During the interview, the researcher will tailor a series of questions to elicit a single type of information from the expert. For example, a researcher may need to ask numerous questions to extract sufficient information that will define all the cues an expert used in building a situation assessment. The interviewer uses the incident as a framework from which to probe decisions, judgments, and problem solving. Thus, the interviewer can elicit more and better information if given the freedom to adapt the process to suit the situation.

The CDM interviewer generally makes four sweeps through the same incident. The first sweep captures the story. The expert relates in his/her own words a particular incident that challenged his/her skills. This helps the interviewer to understand the dynamics of the incident and to determine whether the incident is suitable for further examination.

To gain a sense of the sequence of events and to identify inconsistencies, the second sweep fixes the incident to a timeline. The interviewer works with the expert to identify the time and duration of the incident and to assemble the events in chronological order. At this time, the expert often begins to bring in more specific details about the incident that were not remembered originally.

Table 2-1

Critical Decision Interview Probes

Probe Type	Probe Content
Cues	What were you seeing, hearing, smelling...?
Knowledge	What information did you use in making this decision, and how was it obtained?
Analogues	Were you reminded of any previous experience?
Goals	What were your specific goals at the time?
Options	What other courses of action were considered, or were available to you?
Basis of Choice	How was this option selected/other options rejected? What rule was being followed?
Experience	What specific training or experience was necessary or helpful in making this decision?
Aiding	If the decision was not the best, what training, knowledge, or information could have helped?
Time Pressure	How much time pressure was involved in making this decision? (Scales varied.)
Situation Assessment	Imagine that you were asked to describe the situation to a relief officer at this point, how would you summarize the situation?
Hypotheticals	If a key feature of the situation had been different, what difference would it have made in your decision?

The third sweep employs cognitive probes to detect shifts and elaborations to situation assessments and to identify decision points. The cognitive probes examine goals, cues employed, missing and incomplete cues, expectancies, and courses of action.

The fourth sweep searches for errors, either those committed by the expert or hypothetical errors that might have been committed by people with less experience. In some cases, the expert will talk about mistakes that a novice would make or compare expert performance against that of a journeyman.

Although memories for past experiences cannot be assumed to be perfectly reliable, the method has been successful in eliciting the perceptual cues and details of judgment and decision-making strategies that are generally not captured with traditional reporting methods (Crandall, 1989). Taynor, Crandall, and Wiggins (1987) also showed that the CDM coding had good reliability, both between raters and for the same decision maker at different points in time. Moreover, the CDM provides this information from the perspective of the person performing the task and can be particularly useful in identifying cognitive elements that are central to performance.

The above describes our general model of CDM. However, this direct approach is not the most effective in all situations; it must be modified for each domain. Therefore, the first step in every project in which we anticipate using CDM is to learn about the domain and review the data requirements so that we may tailor the CDM for the specific project.

Our first step in this process was to learn about SAL operations and the B-757/767 aircraft. SAL provided extensive support to accomplish this goal; the researchers engaged in the following activities:

- Observed training sessions in flight simulators
- Conducted tutorial sessions with flight instructors
- Conducted question and answer sessions with various SAL personnel
- Extensively reviewed several SAL publications including the B-757/767 Pilot's Reference Manual, the B-757/767 Operations Manual, the B-757/767 Proficiency Check LOFT Study Guide, and various materials prepared by SAL as part of AQP development

The second adjustment was to the interview process itself. In modifying the CDM, we considered two principal factors: the goals of the project and the nature of the experts' experience.

The interviews with experts had two specific objectives. The first was to elicit sufficient information to accurately describe the flight tasks under investigation. The second was to elicit information about evaluating pilot performance that would serve as a basis for developing the performance measurement instrument. Therefore, we conducted two different interviews with each expert.

The first focused on deriving a detailed description of the tasks performed within the context of the LOFT scenario. The information obtained corresponded to that used by other analysts to develop a task analysis. We solicited detailed information about the performance requirements and the salient events of this scenario. None of the SMEs have flown this specific scenario, but they all had significant line experience in the Cincinnati hub. So, they had no difficulty describing flight operations around Cincinnati. We conducted these interviews in two phases. In the first phase, we asked the SME to provide a synopsis of the major events and crew performance throughout the scenario. The SME then affixed the major performance elements and events to a timeline that began when the crew entered the aircraft. In the second phase, we used the synopsis as a guide for the SME to identify specific performance requirements throughout the flight. The SMEs provided these details for each crew member. Thus, these interviews resulted in a detailed accounting of performance requirements for each crew member relative to a timeline and major events in the scenario.

The second interview focused on eliciting information about how expert instructors evaluate pilot performance. The effective use of the CDM depends on the ability of the experts to draw on their experiences. This produced a unique problem for this study with the abnormal conditions (single-engine operation) under investigation. In addition, none of the interviewees had previous experience with the LOFT scenario that was used in this study. However, all of the experts had considerable past experience with each of the individual flight maneuvers under consideration and with training and evaluating student aviators on these maneuvers. Therefore, we focused the interviews on pilot evaluations. In addition, the experts had difficulty discussing specific instances or flight crews that they had observed. In most cases, they did not have memory of specific incidents. Therefore, to obtain more in-depth information, we asked the experts to consider hypotheticals. That is, we focused the interviews on imaginary crews. We asked the experts to use their experience to describe an outstanding flight crew, the crew that they would like to emulate. The experts constructed incidents around the imagined best and worst crews. From these hypothetical incidents we obtained criteria for distinguishing good and poor performance on the tasks of interest.

2.2.2 Interview Procedures

We conducted interviews with six B-757/767 flight instructors assigned to SAL's Flight Training Department. Their primary duties were to train and evaluate flight crews in flight simulators. Each of the instructors volunteered to participate in the AQP development program; their participation counted as part of their normal work day.

Interviews with each instructor required one full day. In some cases, the instructors were available additional days. This additional time was used to clarify and elaborate points made previously. All interviews were recorded on audio tape and subsequently transcribed verbatim.

2.3 Pretest and Revise PMI

Prior to conducting a complete evaluation of the PMI, we pretested the draft instrument. The objectives of the pretest were to eliminate typographical and technical errors, to verify the rationale for each item on the instrument, and to maximize the instrument's usability.

Five B-757/767 instructor pilots reviewed five video tapes of different flight crews composed of instructors who performed the target LOFT scenario. The instructors used the draft PMI to record and evaluate crew performance on each tape. The instructors completed the forms independently. In addition, the instructors had as much time as they needed to complete the evaluations and could replay the performance of any crew as frequently as desired.

We created the five video tapes used in the pretest by recording three volunteer instructors as they flew the same scenario five times. The crew performed differently during each iteration according to scripts prepared by the researchers. The scripts contained examples of good, average, and poor performance identified during the knowledge elicitation process.

Three of the five instructor pilots had participated in the knowledge elicitation phase of this project; they were familiar with the project's objectives and background. Two of the instructors were new to the project. Prior to conducting the pretest, we briefed all of the participants about the project and about how to use the PMI. We discussed each of the items on the PMI and how the instrument should be used. In addition, we reviewed several hypothetical situations for the instructors to practice using the instrument.

After the instructors completed a rating form for each of the video tapes, they were debriefed by the researchers. During the debriefing, the researchers solicited comments about the appropriateness and rationale for each marker, information that would be available in a flight simulator that was not available on the video tapes, and generally how to improve the PMI.

The researchers revised the draft PMI based on the results of this pretest. Then, we evaluated the draft instrument as described in the following section.

2.4 Evaluate PMI

The objective of this evaluation was to determine whether experienced instructors could evaluate flight crews reliably with this instrument, whether they could use the form with sufficient ease, and whether the target concepts were instilled in the PMI.

Eight volunteer B-757/767 instructors participated in this evaluation. The instructors were matched into four pairs based on their work schedules. Each pair of instructors observed and evaluated performance of four different flight crews as the crews conducted the LOFT sessions required as part of their recurrent training. Thus, we observed a total of 16 different flight crews in training. We made no attempt to select pilots to serve as subjects in this study; we identified pilots based on the training schedule and the volunteer instructors' work schedules.

The subject flight crews were normal line pilots enrolled in the recurrent training course. None of the pilots had met their fellow crew members prior to their recurrent training. In addition, none of the pilots in training had flown the LOFT scenario previously.
A pair of instructors observed each LOFT session. One instructor operated the simulator, provided the necessary radio communications, and supervised the training session. The other instructor acted as an observer. Both instructors completed a PMI as the flight crew progressed through the assigned mission.

After observing four flight crews complete their LOFT, each instructor completed a survey instrument that solicited opinions about the PMI. Appendix B presents the entire survey instrument. The survey included nine items that required instructors to indicate the strength of their agreement with statements about the PMI. The statements concerned the applicability and usability of the PMI. The instructors indicated their opinions on a five-point scale anchored at its extremes to "Strongly Agree" and "Strongly Disagree." In addition, the survey included five open-ended items that solicited opinions about the appropriateness of markers and how to improve the form.

3. FINDINGS

The primary objective of this project was to develop and evaluate a prototype performance measurement instrument that integrated the measurement of CRM skills with technical skills. However, training programs are not designed around performance measurement capabilities; they are designed around thorough descriptions of the tasks and clearly stated proficiency and training objectives. The purpose of performance measurement in AQP is to determine if the student has achieved these training objectives. Therefore, before developing measures of performance, we had to develop task descriptions and proficiency objectives for training.

As we began this project, neither of these documents were available for the B-757/767 aircraft. In fact, none of the AQP applicants had completed the analysis phase of their program development or written proficiency objectives. Thus, we did not have any models for reference. We utilized information elicited during the interviews with experts to build task descriptions and then to develop the proficiency objectives for the tasks investigated in this study.

Each step of the project contributed to building the PMI. Thus, we discuss the findings of each step separately and in the order in which we accomplished them. We present them in the following order:

- Task elements
- Performance standards
- Proficiency objectives
- Performance measurement instrument
- Pretest results
- Evaluation results

The task descriptions and the proficiency objectives were intended only to provide a framework for the PMI. They provide examples of one approach to development, but they address only a small subset of tasks and conditions and are not intended to be exhaustive.

3.1 Task Elements

Our first step in developing a PMI was to identify exactly what a flight crew must accomplish as part of the LOFT scenario. To develop instruments that measure performance, we had to identify the performance requirements first. We referred to SAL's Job List and task analyses for information, but these were not complete. First, the task analysis was not complete for the B-757/767 aircraft. Second, the task analysis produced performance data based on a standard flight profile (preflight, takeoff, climb, cruise, descent, approach, and landing). It did not describe tasks within the context of a specific scenario. Different scenarios may have different performance requirements not captured by the standard flight profile. The task analysis might omit performance elements essential to successful performance in the scenario investigated in this study. Since our focus was to measure performance under very specific conditions in one specific scenario, we relied heavily on information elicited from the SMEs to identify the performance requirements of this scenario. The following is a synopsis of the scenario:

- Prepare for a flight from Cincinnati to Salt Lake City
- Execute a low-visibility takeoff
- Experience a single-engine failure at 200 feet AGL immediately after takeoff
- Return to Cincinnati
- Execute a single-engine Category II ILS approach to Cincinnati
- Execute a Single-Engine Missed Approach when directed by ATC
- Perform a single-engine climbout and cruise to an alternate airport

The Job List identifies several tasks that the crew performs in this scenario. They are listed below:

- Perform flight-planning operations
- Perform preflight/before-start operations
- Perform pushback/start operations
- Perform taxi/before-takeoff operations
- Perform normal static takeoff operations
- Perform after-takeoff/climb operations
- Perform Category II ILS approach operations
- Perform automatic go-around operations
- Perform cruise operations

However, these tasks are not at a sufficiently detailed level to identify performance requirements or standards. We asked the SMEs to provide a step-by-step account of what a flight crew would do as they conducted this scenario, and attach the crew performance to a timeline. That is, the SMEs identified the performance elements of each crew member for the entire scenario and when each element should occur chronologically within the scenario. The SMEs were remarkably similar in their reports. There was very little difference between experts about the performance elements or when they occurred in the scenario. Three of the accounts were virtually identical.

Table 3-1 presents a summary of the SMEs' accounts of the performance requirements in chronological order. This summary includes both crew performance and significant events in the scenario. The table reports what each crew member does during the scenario and when it must be done. The SMEs generated this information assuming that no unusual conditions existed except those incurred as part of the LOFT scenario. The SMEs recognized that many external factors can affect both what a crew does and the timing of the events, but they did not play a role in this study. These factors include such things as weather, equipment malfunctions, traffic, ATC problems, and passenger problems.

Table 3-1

Chronological Order of Events and Crew Performance

Time	Performance
0:00:00	Preflight CP performs walk-around C gets paperwork including flight plan
0:01:00	C begins entering FMS data
0:07:00	CP starts preflight cockpit setup CP verifies FMS entries C begins checklist C gives flight attendant brief
0:10:00	CP gets clearance from ATC C monitors
0:18:00	CP checks flight plan
0:20:00	Gate agent delivers final paperwork including final weight and weather ACARS data and final runway entered into the FMS
0:25:00	Close aircraft door CP reads numbers from Aircraft Weight and Balance Sheet C loads numbers into FMS Ground chief notifies C that aircraft is ready C talks to tug driver CP talks to tower CP requests clearance for pushback C monitors clearance C calls for Pushback/Before Start Checklist
0:27:00	Begin pushback Ground chief calls ready for engine start Flight crew starts engines Set brakes Disconnect tug
0:30:00	C performs in/out salute with ground chief CP requests taxi instructions
0:35:00	C calls for After-Start Checklist Flight crew may start second engine

Table 3-1 continued

0:37:00	C calls for Before-Takeoff Checklist CP performs Before-Takeoff Checklist
0:40:00	CP notifies tower when ready C and CP monitor tower
0:44:00	C or CP moves aircraft into position on runway and hold
0:45:00	Tower clears aircraft for takeoff C calls for heading hold CP selects heading hold Set exterior lights C advances throttles C calls for EPR at 1.1 EPR CP selects EPR Throttles advance to takeoff power C maintains hand on throttles At 80 knots, CP calls out 80 knots, throttle hold, engine instruments check CP calls V_1 and V_r At V_r, C initiates 3 degrees/second rotation CP calls V_2 Aircraft departs ground
0:46:00	C continues to rotate aircraft to 15-18 degrees CP monitors instruments and calls positive rate on barometric altimeter C calls for gear up CP raises gear
0:46:10	At 200 feet AGL, one engine fails C notices that control input required to maintain aircraft control Autothrottle disengages C lowers nose to 12.5 degrees and applies rudder CP monitors instruments CP diagnoses problem and notifies C
0:46:20	At 400 feet AGL, C instructs CP to contact tower to declare an emergency, coordinate a straight out departure, and request a return to Cincinnati C climbs out straight and maintains aircraft control C may initiate memory items on Engine Fire/Failure Checklist

Table 3-1 continued

0:47:00	At 1000 feet AGL, C lowers nose to 7.5 degrees to accelerate C calls for Altitude Hold CP selects Altitude Hold C tracks command bar to accelerate C calls for CP to raise flaps on schedule At $V_{ref}30 + 40$ C calls for Flight Level Change on MCP C calls for Continuous Power CP sets Continuous Power
0:49:00	C directs CP call ATC for clearance for approach to Cincinnati C initiates proper climb and turn C calls for Engine Fire/Failure and After Takeoff Checklists CP performs checklists C engages autopilot
0:51:00	On downwind, C calls for Descent Checklist C possibly informs company, flight attendants, and passengers C instructs CP to review the approach, tune and identify the radios, set altimeters and airspeed bugs, and program the FMS CP notifies C when complete C transfers control to CP by verbalizing aircraft status on MCP CP acknowledges transfer C reviews and briefs the approach, checks the FMS, and cross checks the settings of the altimeters and airspeed bugs C retakes control of the aircraft using the MCP to brief aircraft status C directs CP to coordinate a straight out missed approach C calls for Approach Checklist and CP complies
0:54:00	ATC provides turns to final approach Slow to Flaps 5 speed C sets speed on MCP and calls for Flaps 5 ATC clears aircraft for final approach
	C selects Approach mode on MCP CP monitors flight instruments and notifies C of first movement and capture of localizer and glide slope With glide slope alive, C calls for gear down, Flaps 20, and update speed C calls for Before Landing Checklist
0:56:00	CP notifies tower when the aircraft is at the outer marker At 1500 feet AGL, CP verifies verbally that the scorecard shows Land 2.

Table 3-1 continued

0:57:00	ATC directs aircraft to execute a missed approach C calls missed approach C activates go-around mode and advances throttles to go-around power C calls for Flaps 5 CP selects Flaps 5 CP monitors flight instruments and calls positive rate on barometric altimeter C calls for gear up CP raises gear C directs CP to notify tower they are going around
0:58:00	At 1000 feet AGL, C directs CP to select Altitude Hold C directs CP to retract flaps on schedule C directs CP to select Flight Level Change and Continuous Power C calls for After Takeoff Checklist C directs CP to coordinate with company and ATC to identify an alternate airport and obtain clearance

Table 3-1 presents a linear description of crew performance for the entire scenario. Our concern was to ensure that we identified all of the important elements of crew performance for this limited set of conditions. Therefore, we did not begin by attempting to decompose crew performance into flight phases, tasks, subtasks, and elements as seen in a typical behavioral task analysis. This approach ensured that we identified the important elements of behavior without the constraints of categorizations. Subsequently, we classified the performance elements into the tasks identified in SAL's Job List. However, as we discuss later, this was not altogether successful for two reasons. First, some elements of performance, particularly cognitive performance, extend across more than one task. Second, the current Job List does not account for all of the performance elements identified in this study.

These data subsequently served as a guide for identifying the performance standards, constructing the performance objectives, and developing the PMI. We used this basic information to focus subsequent interviews with the SMEs to elicit information about indicators of good and poor performance.

3.2 Performance Standards

The next step was to identify standards of proficient performance for this scenario. Our CDM interviews elicited information about the cues and markers that allow the SMEs to distinguish excellent performance from minimally acceptable performance. During the interviews, the SMEs anchored their discussions to the specific situation of the scenario. They described how excellent pilots would perform this scenario and the mistakes that minimally acceptable pilots would make.

Our discussions about pilot evaluation stirred some rather intense responses from all of the SMEs that we interviewed. Initially, we attempted to establish boundaries for the lower end of performance by determining what constituted unsatisfactory performance for pilots in recurrent training. Invariably, these questions raised a lot of concern. Failures rarely occur during recurrent training.

It was difficult to find fixed criteria that evaluators use to fail a pilot or crew on a recurrent checkride. They claim there is no single criterion, even though the manuals or standards may designate certain performance levels as unsatisfactory. The evaluators are extremely reluctant to fail someone. This might reflect the atmosphere that recurrent training is to be an open and non-threatening environment for learning and polishing skills. Every pilot in recurrent training has already qualified. Thus, it is not perceived to be a time that pilots have to defend their flight proficiency or show why they should be pilots. They have already passed that stage in their careers. However, recurrent training prevents a mixed environment in which pilots must pass a checkride before returning to the line.

All of the recurrent students are extremely experienced professional pilots. The evaluators argued that it would take unsatisfactory performance on several dimensions (not just technical) for them to fail someone. For example, the pilot would have to display unsafe technical skills coupled with a poor attitude reflected by her/his lack of study, demeanor, dress, or concern.

Many of the evaluators expressed the opinion that pilot differences are often a matter of technique. Simply because a student performed a task in a manner that differed from the way the evaluator would do it, does not mean the student was wrong. Evaluators claim that the student may prefer a different, but equally safe and effective technique. Thus, they do fail student pilots or criticize them for doing something differently.

Another problem is that most of the proficiency check airmen are not regular line pilots, because they are and have been assigned to the flight-training department for a long time. In fact, they fly the airplane and the simulator infrequently. Thus, they do not have the accumulated operational experience that their students have. Simulator instructors may have some reluctance to criticize a student pilot because the student may, in fact, have a better way of doing business.

After bypassing the hurdle of describing unsatisfactory performance, the SMEs could quite adequately discuss distinctions between their model crews and those that barely meet the criteria for success. We focused on the three principal tasks included in this leg of the scenario: Takeoff, Category II ILS Approach, and Missed Approach. There were a couple of exceptions. All of the SMEs discussed the value of the Departure Briefing to subsequent events during the flight. The brief sets the tone for the flight and covers much information that is relevant as the crew flies the mission. Therefore, we studied the criteria for evaluating the quality of the Departure Brief.

The SMEs identified several criteria for distinguishing superior crews that do not occur during one of the three principal tasks. These criteria occur simply as an artifact of this particular scenario. There are several performance markers that the SMEs look for as the crew transitions from the Takeoff to the ILS Approach. However, the markers do not occur within the strict confines of either task as they are described in SAL's task analysis.

The SMEs reported considering many technical and CRM-related behaviors when assessing pilot and crew performance. Some were discussed by all SMEs; others were not confirmed by the majority. The following discussion concerns only those behaviors that were mentioned by the majority of the SMEs interviewed. The discussion presented here augments the description of the performance markers presented later.

During normal takeoff, the SMEs considered few technical markers. These included rotation rate and airspeed maintained by the pilot flying. Typically, the automation handles the remainder of the flying tasks. The relevant CRM factors include completing the callouts required by procedures and the briefings. However, several factors become more critical as the engine fails after takeoff. First, the pilot flying (PF) must maintain aircraft control and adjust the airspeed. Team factors also become more important. Both pilots must detect the problem and notify the other. In addition, the pilot not flying (PNF) should notify the PF when specific checklists are complete. Normal procedures require that the crew reconfigure the aircraft for climb as they pass through 1000 feet AGL. This altitude comes quickly, and the crew may not be ready to begin retracting the flaps. They may still be occupied trying to control the aircraft, declaring an emergency, or diagnosing the problem. The crew should not attempt do too many things at once. Their first priority is to fly. Once the aircraft is under control, then they can move on to the other tasks.

The SMEs indicated that good crews will typically reconfigure the aircraft to a traffic pattern flap setting (Flaps 5) that would allow them to remain at a lower airspeed as they prepared for the approach. Good crews immediately recognize that they will return for an approach to Cincinnati and that the lower airspeed will provide sufficient time to prepare. Less proficient crews do not immediately recognize the ramifications of selecting a higher airspeed that will be manifested later. Crews that go to clean speed tend to be rushed when preparing for the approach or run out of time completely and have to truncate their preparation.

The SMEs described several performance criteria that should occur after the takeoff and before the crew initiates the approach. The crew should communicate with the passengers, complete two checklists, prepare for the approach, and fly the airplane. Communicating with the passengers should occur only if the crew has sufficient time. The crew should verify the weather conditions for an approach and identify the correct approach to be flown (i.e., Category II). The SMEs indicated that the PF's flying technique during this segment provides some insight into his/her ability. At some point, the PF should engage the autopilot.

However, during a turn to downwind may not be the best time to do it. The SMEs indicated that the better pilots wait until the aircraft is established on downwind before engaging the autopilot; in this scenario, better pilots hand fly the airplane through the turn to final. The reason is that the PF must manipulate the rudder trim through the turn and that the autopilot is sensitive to any out of trim conditions. The autopilot will not engage with an out of trim condition and often disengages in this type of maneuver. This results in a loud alarm in the cockpit and two crew members distracted by a relatively minor problem. The better pilots recognize this and avoid the distractions by delaying the autopilot until after the turn.

The Category II ILS Approach provides numerous opportunities to evaluate a crew's CRM skills. However, the crew should engage the autopilot for the approach, so there are very few technical skills for the human pilot. The approach contains numerous mandatory CRM tasks such as required callouts, checklists, and division of duties. Good crews comply with these; poor crews do not. The approach also has numerous other opportunities for good crews to distinguish themselves by displaying their team skills.

For example, the crew can maintain a shared awareness of the situation through good briefings and by using the Mode Control Panel (MCP) when they exchange controls. The PF uses the MCP settings to describe the status of the airplane and his/her intentions as they pass control of the airplane. Workload distribution is another example. Approaches contain periods of time with little to do and periods that have too much to do. Good crews recognize this and distribute their tasks evenly across the approach; they do as much during the downwind as possible. Another clue to good performance is whether the crew acknowledge that they do not have sufficient time to complete their preparation and request additional time from ATC. The crew may make one mistake and fall behind. But, they should not compound that mistake by attempting to execute the approach before they are prepared. Finally, good crews remain cognizant of what the automation is doing and the details of the approach, even though neither pilot is hand flying.

The Missed Approach contains several critical indicators of technical and team performance. Line crews rarely perform missed approaches, and this one is compounded by the inoperative engine. Therefore, line crews often make technical errors in attempting a single-engine missed approach. For example, some pilots do not initiate the go-around correctly. They may not depress the switch sufficiently to activate the go-around mode. They may not remember that autothrottles do not function during single-engine operation and fail to advance the throttles. Another common technical problem is that the PF has difficulty controlling the airplane as s/he applies go-around power to one engine. Team performance plays a role in the missed approach as well. The PNF must ensure the completion of the required callouts and support the PF as needed. Again, the crew should ensure that the aircraft is under control before they begin to reconfigure the aircraft.

The diversion to an alternate airport requires team coordination. The pilots should be proficient in the technical tasks of this segment since they closely approximate normal line flying. However, both crew members should be involved in the decision to divert and they should consider all relevant factors. Finally, the good crews reduce their workload by using the navigation features of the Flight Management System and refer to the single-engine operation information.

Our interviews with the SMEs helped to crystallize an image of the outstanding crew that they use as a model. This crew is quite competent technically. They do not get flustered and handle the aircraft quite smoothly. During an engine failure, an observer would not notice anything different. The PF would apply the correct control inputs with no heading or roll excursions. The crew always has the correct procedure close at hand. Perhaps more important, the outstanding crew is a thinking crew as well as a good team. The crew is always thinking ahead. They anticipate what the aircraft is going to do. They understand the ramifications of their actions on future events. The crew plan well; they examine conditions and pre-plan responses to events that they can anticipate. Finally, this outstanding crew do not leave the automation in charge and unattended. The crew monitor the automation and the flight controls to maintain an awareness of the situation. They remain in the loop and prepared to resume control when desired.

3.3 Proficiency Objectives

From the task descriptions and the performance criteria we developed a Terminal Proficiency Objective (TPO) for each task and a Supporting Proficiency Objective (SPO) for each subtask. TPOs are proficiency objectives for tasks; SPOs are proficiency objectives for subtasks. Each includes three components: a performance statement that describes the task or subtask; a conditions statement that describes existing environmental, crew, and equipment conditions; and a standards statement that defines proficient performance.

The AQP advisory circular describes a task as comprising a number of subtasks required to accomplish the task. We chose to follow this description and to focus pilot performance measurement at the subtask level. Thus, we wrote TPOs to reflect the specific subtasks subsumed under each task. SPOs contain the specific statements of performance and standards.

One of the critical issues for drafting proficiency objectives is to determine how many separate SPOs are needed for each task. One could write a separate SPO for each subtask and condition set. However, this would soon become unwieldy.

We chose to create a separate SPO for each subtask and for each condition set only if the condition set significantly altered either the performance statement or the performance standards. For example, we drafted an SPO for the subtask "Departure Briefing." One SPO should be sufficient for all condition sets because varying conditions do not alter basic content or performance of the briefing.

ILS Approach provides another example of this method for constructing proficiency objectives. Many task analyses separate ILS Approaches into Category I, II, and III. After reviewing these three types of approaches, we concluded that the existing condition sets did not sufficiently alter performance to justify additional proficiency objectives for Category I and II approaches. The reduced visibility affected the decision heights, but does not alter the basic performance of the task. For example, Category III approaches differ from Category I in that only the captain can perform them and decision heights become alert heights with no visual references required.

Therefore, we drafted TPOs and SPOs for each of the tasks, subtasks, and conditions included in the LOFT scenario. We drafted SPOs for each of the relevant subtasks under normal conditions as well as the single-engine operation. This provides a basis for comparing objectives for the same subtask under different environmental or equipment conditions. Table 3-2 presents a list of the TPOs and SPOs drafted for this project.
Appendix C includes all of the TPOs and SPOs drafted.

3.4 Performance Measurement Instrument

We used the Pilot Performance Description Report (PPDR) as a framework for developing the PMI in the project. Smith, Flexman, and Houston (1952) developed the PPDR to reduce the subjectivity present in evaluations of pilot performance and to provide a method for standardizing flight evaluations. Greer, Smith, and Hatfield (1962) and Prophet and Jolley (1969) later modified the PPDR. The PPDR comprises two sections that must be completed by the evaluator as the pilot performs the maneuver. The first section provides a means of describing how the pilot performed the task. It contains scales anchored to performance standards on which the evaluator describes specific aspects of performance. For example, descriptive scales for a certain task may include measures of airspeed, altitude, or pitch attitude. The second section of the PPDR is a subjective rating scale on which the evaluator provides an assessment of overall performance on the tasks. The descriptive scales and the performance rating scale vary depending on the task and the user's objectives.

Greer et al. demonstrated that overall proficiency ratings with PPDR descriptive scales are more reliable than ratings made without the PPDR. Versions of the PPDR have been used effectively to evaluate aviator performance in a variety of investigations (Childs, Spears, & Prophet, 1983; Kaempf & Blackwell, 1990; Shelnutt, Spears, & Prophet, 1981).

Table 3-2

Proficiency Objectives for Each Task and Subtask.

Terminal Objectives	Supporting Objectives
Preflight/Before Start	Departure Briefing (all conditions)
Takeoff	Assess Performance and Environmental Factors (all conditions)
	Takeoff Roll (normal)
	Takeoff Rotation and Liftoff (normal)
	Initial Climb (normal)
	Initial Climb (single engine)
Cat. II ILS Approach	Approach Preparation (normal)
	Approach Transition (normal)
	Final Approach (normal)
	Final Approach (single engine)
Missed Approach	Initiation (normal)
	Initiation (single engine)
	Initial Climb (normal)
	Initial Climb (single engine)

Using the PPDR as a framework, we needed to identify the components of performance that are critical for assessments. These would provide the descriptive measures of performance.

We used data obtained during our interviews with SMEs to develop these descriptive measures of performance for the selected tasks. The SMEs also used these interviews as opportunities to ensure that the researchers understood the complexities of assessing pilot performance in a major airline. The topics included:

- Skill level of the pilots in training
- Criteria currently used to evaluate pilot performance
- Workload of evaluators conducting the training sessions.

These issues shaped our approach to developing the instrument. The following paragraphs discuss some of the more pertinent issues. Then we describe the PMI itself.

3.4.1 Assessment of Pilot Performance

The focus on pilots in recurrent training affected how we chose to measure their performance. These pilots are not training to achieve a rating; they are fully qualified and licensed line pilots. Therefore, they are quite proficient on the majority of flying tasks. They maintain this proficiency by conducting their normal operational flying duties.

Recurrent training is designed to provide an opportunity for line pilots to refresh their skills on those tasks that they do not perform habitually on the line. Since the pilots have not performed some tasks since they last attended training, their skills have probably degraded to some degree. These tasks include primarily abnormal procedures and unusual environmental conditions. Usually, line pilots require minimal, if any, training to regain a proficient level of performance.

The result of this is that we expected to see very few unsatisfactory performances during recurrent training sessions. The great majority of tasks would be completed satisfactorily. SAL's training records verified that this is the case. Line pilots rarely repeat maneuvers during training to bolster sub-standard performance. In fact, maneuvers judged the most difficult required repeats on less than ten percent of the trials attempted.

We expected to see considerable variability among pilots who were proficient on the tasks. Some pilots would be better than others, but in most cases, they would all be satisfactory. Therefore, we elected to construct performance measures that would be sensitive to variability in proficient performance rather than variability in sub-standard performance. The scales would provide adequate space for the evaluator to indicate unsatisfactory performance when it occurred, but the bulk of the scales would attempt to discriminate varying degrees of performance that exceeded standards.

Another issue brought to our attention was the workload of the instructor who conducts the recurrent LOFT sessions. We intended to design the PMI to be implemented by the instructor during the LOFT session. Various instructors discussed the heavy workload that these instructors have during certain periods of each scenario. During each session, the instructor serves several roles. Most important, they must observe and evaluate the crew's performance. But, they also must act as other players in the simulation by creating communications from ATC, the company, and others. Finally, they operate the simulator. Any additional work requirements may degrade their abilities to conduct the scenarios effectively. Thus, we had to design the PMI in a fashion that would minimize the distractions and the workload for the instructors.

A final point concerns the importance of expertise in evaluating flying skills. For many years, aviators have considered flying to be a mixture of art and science. There are subjective as well as objective components of flying. Seat-of-the-pants flying is an excellent example. Pilots develop an implicit feel for the airplane that they have difficulty describing. This extends to teaching someone to fly and evaluating their performance. An instructor can read the instruments to judge performance, but their own senses and experience tell them more about the student's performance. These perceptions develop with experience. Objective measures of pilot performance help to standardize training and evaluation. However, we are not yet ready to totally remove the subjective opinion of the expert from the evaluation process. We believe that we still have a considerable amount to learn from the expert evaluator. Thus, we intend to retain subjective measures of pilot performance, but in a framework that will allow us to learn more about the dimensions and features that the expert considers in making these evaluations.

3.4.2 PMI

Figures 3-1 and 3-2 present two versions of the PMI. The original version (Figure 2) was the subject of the pretest. We revised this version based on the results of the pretest, and conducted the formal evaluation using the revised version of the instrument. The following paragraphs describe the revised version of the PMI.

We followed several guidelines in developing the PMI. These guidelines are discussed below. First, to improve instructors' acceptance and reduce their workload, we took considerable effort to keep the form simple. We limited the form to one page and used color to enhance readability inside a darkened simulator. We took seriously the SMEs' complaints about workload. High instructor workload would only serve to reduce their ability to teach students and to evaluate performance. So, every effort was made to reduce their burden.

Second, the instrument is organized around the major tasks performed during this particular flight leg: Takeoff, Category II ILS Approach, and Missed Approach. However, these three tasks, as defined by a task analysis of the B-757/767 aircraft, did not capture all of the critical elements of performance described by the SMEs in this study. A full 25% of the content of the PMI came from activities the job task listing did not address. Thus, we expanded the instrument to include a fourth major task that we called "Enroute." We do not propose that "Enroute" survive intact. However, we included it as a catch-all task to ensure that the PMI reflected all behavioral elements that evaluators used to assess crew performance. Enroute includes two unrelated flight segments. The essential behavioral elements included those that occur after completing the Takeoff and before initiating the ILS Approach. We called this sequence "Turn to Final." In addition, we wanted to study several behavioral elements that occur after the Missed Approach. These involved the processes of selecting an alternate airport and rerouting to that alternate. We called this segment "Alternate."

Another segment that is not subsumed under the appropriate task is the Departure Briefing. On the PMI, we have included this subtask as part of the Takeoff. This briefing is typically considered part of the task Preflight/Before-Start Operations. All of the SMEs interviewed discussed the importance of this briefing for setting the tone of the crew and for identifying information relevant for the remainder of the flight. Thus, we wanted to include it as a critical subtask. However, the remainder of the preflight tasks were not of interest in that they involve primarily following checklists to configure the aircraft for flight. We chose to include the Departure Brief as part of the task Takeoff simply to avoid creating a separate block on the form for this one small segment.

TAKEOFF

Departure Brief
☐ Complete
☐ Eye Contact
☐ PNF Contribute

☐ Complete TO Brief
☐ Complete All Callout
☐ Maintain Centerline
☐ PNF Monitors Inst.
☐ Rotation Rate
☐ Pitch Attitude

☐ PF Announces Problem
☐ PNF Confirms Problem
☐ Heading Control
☐ Roll Control
☐ Adjust Pitch Attitude
☐ PNF Supports PF
☐ Under Control @ 1000'
☐ Cleanup @ 1000'
☐ Use Altitude Hold
☐ Callouts For Flaps
☐ Go To Flaps 5

CREW PERFORMANCE
UNSAT. | SATISFACTORY: Accept. Avg. Expert

CAT. 2 ILS APPROACH

☐ Descent Checklist
☐ Engage AP
☐ Check Weather
☐ Determine App. Cat.
☐ MCP To Exchange
☐ PNF Calls Complete
☐ Verify Settings
☐ Approach Briefing
☐ Coordinate Miss.
☐ Review Airport Chart

☐ Work. Distribution
☐ Comm: Pax, Co.
☐ Ample Time
☐ Request More Time
☐ PNF Monitors
☐ PNF Callouts
☐ Configured at FAF
☐ PF Monitor Throttles
☐ PF Monitor Controls
☐ PF Monitor Trim

CREW PERFORMANCE
UNSAT. | SATISFACTORY: Accept. Avg. Expert

MISSED APPROACH

☐ PF Announce GA
☐ Intend To Use AP
☐ Engage TOGA
☐ Apply Power
☐ Hand Fly GA
☐ Heading Control
☐ Roll Control
☐ Trim

☐ Flaps 5
☐ Complete Callouts
☐ Clean Up @ 1000'
☐ Altitude Hold
☐ A/C Control @ Cleanup
☐ Anticipate Trim Require.
☐ PNF Monitors
☐ PNF Supports

CREW PERFORMANCE
UNSAT. | SATISFACTORY: Accept. Avg. Expert

ENROUTE

TURN TO DOWNWIND
Decision To Return
☐ Timely
☐ Discuss With FO
☐ Consider Weather
☐ Announced

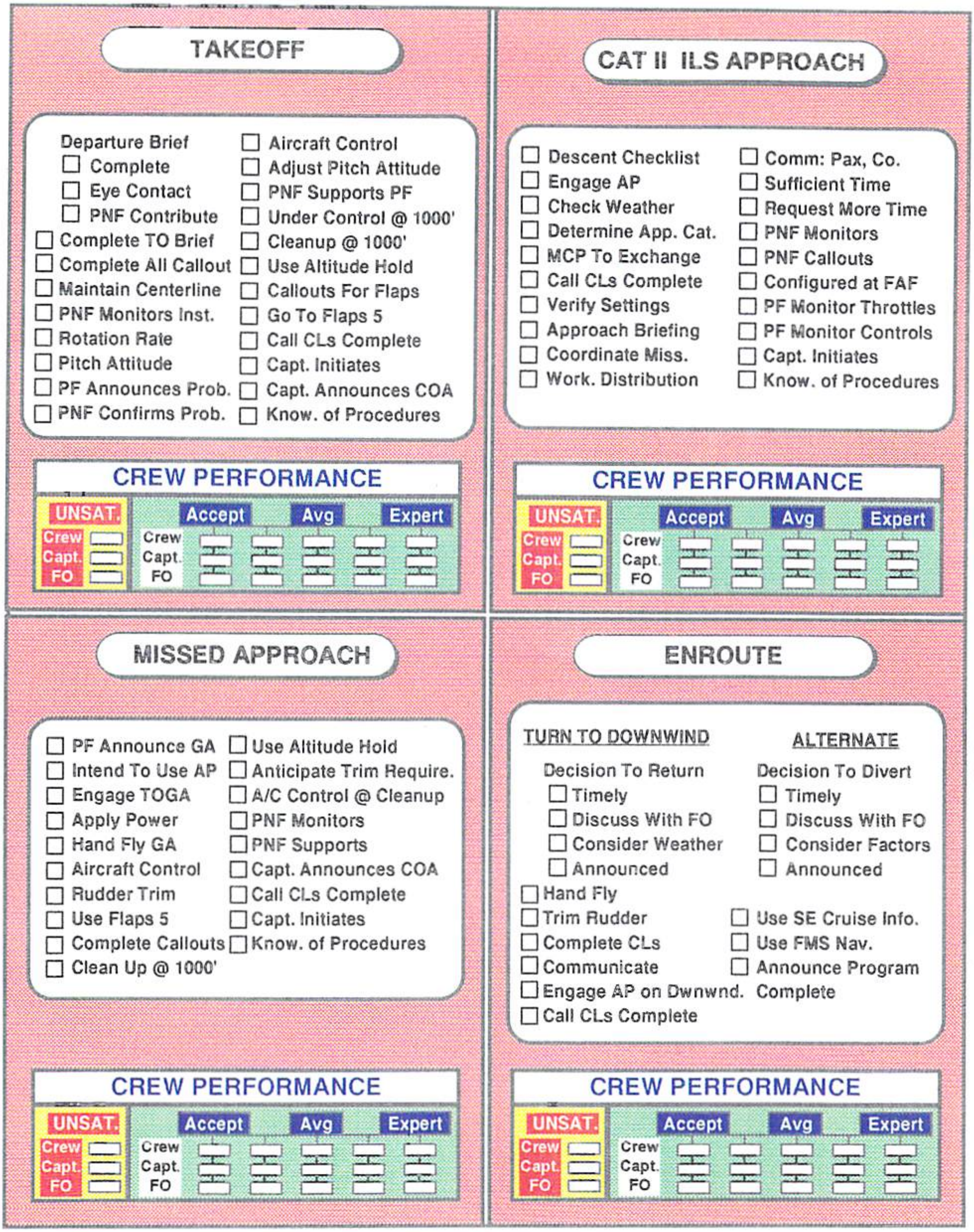
ALTERNATE
Decision To Divert
☐ Timely
☐ Discuss With FO
☐ Consider Factors
☐ Announced

☐ Hand Fly
☐ Trim Rudder
☐ Complete CLs
☐ Communicate
☐ Engage AP on Dwnwnd.

☐ Use SE Cruise Info.
☐ Use FMS Nav.
☐ Announce Program Complete

CREW PERFORMANCE
UNSAT. | SATISFACTORY: Accept. Avg. Expert

Figure 3-1. Original Performance Measurement Instrument



TAKEOFF

Departure Brief
- ☐ Complete
- ☐ Eye Contact
- ☐ PNF Contribute

☐ Complete TO Brief
☐ Complete All Callout
☐ Maintain Centerline
☐ PNF Monitors Inst.
☐ Rotation Rate
☐ Pitch Attitude
☐ PF Announces Prob.
☐ PNF Confirms Prob.
☐ Aircraft Control
☐ Adjust Pitch Attitude
☐ PNF Supports PF
☐ Under Control @ 1000'
☐ Cleanup @ 1000'
☐ Use Altitude Hold
☐ Callouts For Flaps
☐ Go To Flaps 5
☐ Call CLs Complete
☐ Capt. Initiates
☐ Capt. Announces COA
☐ Know. of Procedures

CREW PERFORMANCE
UNSAT. — Crew, Capt., FO
Accept — Avg — Expert — Crew, Capt., FO

CAT II ILS APPROACH

☐ Descent Checklist
☐ Engage AP
☐ Check Weather
☐ Determine App. Cat.
☐ MCP To Exchange
☐ Call CLs Complete
☐ Verify Settings
☐ Approach Briefing
☐ Coordinate Miss.
☐ Work. Distribution
☐ Comm: Pax, Co.
☐ Sufficient Time
☐ Request More Time
☐ PNF Monitors
☐ PNF Callouts
☐ Configured at FAF
☐ PF Monitor Throttles
☐ PF Monitor Controls
☐ Capt. Initiates
☐ Know. of Procedures

CREW PERFORMANCE
UNSAT. — Crew, Capt., FO
Accept — Avg — Expert — Crew, Capt., FO

MISSED APPROACH

☐ PF Announce GA
☐ Intend To Use AP
☐ Engage TOGA
☐ Apply Power
☐ Hand Fly GA
☐ Aircraft Control
☐ Rudder Trim
☐ Use Flaps 5
☐ Complete Callouts
☐ Clean Up @ 1000'
☐ Use Altitude Hold
☐ Anticipate Trim Require.
☐ A/C Control @ Cleanup
☐ PNF Monitors
☐ PNF Supports
☐ Capt. Announces COA
☐ Call CLs Complete
☐ Capt. Initiates
☐ Know. of Procedures

CREW PERFORMANCE
UNSAT. — Crew, Capt., FO
Accept — Avg — Expert — Crew, Capt., FO

ENROUTE

TURN TO DOWNWIND

Decision To Return
- ☐ Timely
- ☐ Discuss With FO
- ☐ Consider Weather
- ☐ Announced

ALTERNATE

Decision To Divert
- ☐ Timely
- ☐ Discuss With FO
- ☐ Consider Factors
- ☐ Announced

☐ Hand Fly
☐ Trim Rudder
☐ Complete CLs
☐ Communicate
☐ Engage AP on Dwnwnd.
☐ Call CLs Complete
☐ Use SE Cruise Info.
☐ Use FMS Nav.
☐ Announce Program Complete

CREW PERFORMANCE
UNSAT. — Crew, Capt., FO
Accept — Avg — Expert — Crew, Capt., FO

Figure 3-2. Revised Performance Measurement Instrument

Finally, we employed two approaches to performance measurement for each task. We instructed the evaluators to wear two hats for us. Under one hat, they served as a data recorder. They simply observed each crew and recorded what the crew did. For each task, the form presents a list of critical performance markers. The evaluator simply checks which one s/he observes. Under the other hat, the evaluators assessed performance for each crew member individually and for the crew as a team. They used their expertise to evaluate how well each crew performed each task. The evaluators provided these assessments using a six-point scale that allowed them to indicate unsatisfactory performance as well as five degrees of acceptable performance.

This two-pronged approach has worked well in other studies (Kaempf & Blackwell, 1990; Kaempf, Cross, & Blackwell, 1989) and provides several benefits. First of all, it allowed us to capture the expertise of the instructors. This is important in that it allows us access to information not available through computers and indicators. Second, the performance markers provide diagnostic information to evaluate the student's strengths and weaknesses, to identify the components of expert evaluations, and to evaluate the effectiveness of the training program. Analysis of the performance markers will reveal specific skills that a pilot needs to work on. However, the markers can also serve more far-reaching purposes. First, they can be used as a means of understanding expert evaluations. Subsequent applications of this approach can identify the components of expert evaluations by studying the relationships between the performance markers and performance ratings. Finally, the performance markers can provide diagnostic information for trouble-shooting the training program. Such analyses can help identify training modules that require revision.

Essentially, the PMI measures performance at two levels, TPO and SPO. The evaluators provide their subjective assessment for performance on the whole task, at the TPO level. In contrast, the performance markers are anchored to the performance statements and standards of the SPOs. The evaluator can use these markers as aids in deriving a subjective rating. However, we do not provide any guidance about how s/he might do this. That is, we do not provide any formula or weightings for deriving a rating of overall performance from the performance markers.

Many of the performance markers are self-explanatory and require no further description. Descriptions of a few are provided below. All of the markers, the rating scales, and instructions for using the PMI are included in Appendix D. The reader should refer to this appendix when reviewing the PMI. We distributed this document to the evaluators to serve as a training tool and as a reference as they worked with the PMI.

During the Takeoff and the Missed Approach, the flight crew can become occupied trying to maintain control of the aircraft and manage the situation. Although procedures call for the crew to begin reconfiguring the aircraft as they pass through 1000 feet AGL, there is no safety requirement for the crew to do so. The crew should retract the flaps when they have time to do it and when they feel comfortable assuming the additional task. Therefore, two performers markers for these two tasks address this. The evaluators can record if the crew has the aircraft under control at 1000 feet AGL and whether the crew begins to reconfigure the aircraft at that altitude. Good performance is characterized by being under control and cleaning up at 1000 feet. Poor performance is characterized by trying to clean up before having the aircraft under control.

Another marker on the takeoff concerns whether the crew retracts the flaps to Flaps 5 or retracts them completely. Technically, the crew may select either position and the accompanying airspeed. This performance marker and several listed under the approach are to determine if the crew is thinking ahead, whether the pilots are ahead of the aircraft. Crews may reactively accelerate to clean speed without considering that this action will abbreviate their traffic and allow less time for the approach. The SMEs indicated that better crews will quickly decide to stay in the traffic pattern and return for an approach to Cincinnati. Thus, they select the traffic pattern airspeed and call for Flaps 5. This allows them to have sufficient time to prepare for the approach. Other crews will not realize the result of excessive airspeed until they reach the end of the downwind leg, receive clearance for the turn, and not be ready for the approach. The higher airspeed did not provide sufficient time for them to prepare, but they did not realize the problem until too late to resolve it. This crew should not compound this problem by making another. If the crew is not ready for the approach, they should not accept. The crew might coordinate an extended traffic pattern with ATC until they are ready and comfortable.

Another indication that a crew is thinking ahead is if they coordinate a straight out missed approach sometime during the Category II ILS Approach. With a single engine, the good crew anticipates that a straight missed approach would be easier.

In addition, the approach contains two markers that the crew remains engaged in the task, maintains a situation assessment, and shares that assessment between crew members. When exchanging controls, the crew can use the settings of the MCP to brief each about the current status of the airplane and what their immediate intentions are. For example, they may cover current altitude, airspeed, heading, clearances, and a reminder of unusual conditions such as the unavailability of autothrottles due to single-engine operation. The SMEs indicated that it is important for the crew to maintain an awareness of the autopilot's functioning and activities. The crew should never simply be along for the ride. On the approach, the PF can monitor the autopilot by following through on the flight controls and by monitoring the flight instruments. The evaluator can detect this by determining if the PF maintains his hands on the flight controls.

Line pilots often experience difficulties initiating a missed approach during the LOFT sessions. They make two common errors. First, they may not remember that the autothrottles are inoperative and fail to advance the throttles manually. Second, they may not activate the Takeoff/Go-Around button on the throttles. Either error often leads to similar results. The PF recognizes that the automation is not doing what s/he wants and disengages the automation to hand fly the missed approach. It is not a technical error to hand fly a missed approach; the crew can choose to use either method. However, the crew are not using all of their resources if they choose to hand fly the missed approach. Furthermore, the crew commit a technical error if they intend to use the automation but fail to engage it properly. Therefore, the PMI contains two performance markers, "Intend to use AP" and "Hand fly GA." The evaluators can use these two blocks to indicate the crew experienced the types of problems described above.

The evaluators were instructed to mark the boxes next to the performance markers that they observed as the crew preformed the task. After the crew completed the task, the evaluators then provided an assessment of the crew's overall performance on that task. We instructed the evaluators to use their own judgment in making these assessments, to rely on their past experiences and familiarity with the line flight crews. They could refer to the performance markers if desired. The instructors made two judgments for each task. First, they decided if the crew performed the task satisfactorily or not. If not, then they marked the "UNSAT" box and progressed to the next task. If performance was satisfactory, then they used a five-point scale to rate the crew or pilot relative to the average SAL crew. Complete descriptions of the verbal anchors are presented in Appendix D.

3.5 Pretest Results

Video tapes proved to be an effective medium for pretesting the PMI. However, the video tapes have several limitations that preclude their use for a thorough evaluation. First, the evaluators could not observe many crew behaviors on the video tapes. The position of the camera and the low light levels limit what the viewer can see. None of the aircraft instruments are visible. Only gross movements of the hands and head are clearly visible. Thus, it is difficult to determine the quality or appropriateness of control inputs or where the pilots focus their attention.

In light of these limitations, the pretest produced good information about revisions for the PMI. These came in the form of open-ended responses from the instructors. Generally, the revisions pertained to adding, deleting, or altering specific performance markers. The instructors expressed beliefs that the approach was sound and would facilitate evaluations of flight crews. The most significant revision made to the PMI was the inclusion of rating scales to assess the performance of each crew member individually. This suggestion was made to fulfill an AQP requirement to evaluate the performance of individual crew members in each curriculum.

Results of the pretest indicated that the instructors had little difficulty using the performance rating scale. Again, they only provided assessments for the crew. The mean ratings awarded by each instructor varied from 2.55 (S.D.=1.9) for one instructor to 3.45 (S.D.=2.1) for another. Inter-rater reliability proved to be remarkably high. Table 3-3 shows the correlations between the ratings awarded by the five instructors; all correlations were significant ($p<.01$). These results were encouraging in that they indicated the instructors were standardized in their perceptions of crew proficiency and they could reliably use the PMI

Table 3-3

Correlation Provided by Each Instructor During the Pretest

INSTRUCTOR	INSTRUCTOR 1	2	3	4	5
1	-				
2	.948	-			
3	.891	.927	-		
4	.964	.948	.916	-	
5	.936	.910	.803	.924	-

3.6 Evaluation Results

We used three types of measures to evaluate the PMI: evaluator responses to the performance markers, subjective performance ratings awarded by the evaluators, and the evaluators responses to the items in the survey instrument. Each of these is discussed below.

To review how we conducted the evaluation, a total of eight instructors employed the PMI as they observed line crews conduct the LOFT portion of their recurrent training curriculum. Instructors were assigned to pairs; each pair observed four crews perform the mission. No two pairs of instructors observed the same crew in training. Thus, to assess inter-rater reliability we must look at the reliability within each pair of instructors.

3.6.1 Performance Markers

Observing and recording the performance markers demanded most of the effort from the evaluators. They had to accomplish several tasks simultaneously while trying to detect some rather subtle behaviors. It was of interest to determine if the evaluators had difficulty detecting the behaviors and if they could do so reliably.

We assessed reliability between evaluators by calculating the percent agreement between evaluators in each pair and calculating the average of all four pairs. The four tasks comprised varying numbers of performance markers. Takeoff included 22 markers, CAT II ILS Approach included 20, Missed Approach included 19, Enroute included 18. Thus, the pairs of evaluators had a total of 79 opportunities to agree for each crew and 316 opportunities across all four crews. The percent agreement measure simply divides the number of times the pair agreed by the number of possible opportunities.

The evaluators had little difficulty indicating which performance markers they observed and which they did not. The average percent agreement for the four pairs of instructors was 77.2%, 77.5%, 86.1%, and 92.7%. Clearly, the evaluators demonstrated a rather high level of reliability despite the brief training they received prior to using the PMI.

In addition, we looked at the inter-rater reliability for each task separately. The average percent agreement for the four pairs of evaluators was remarkably similar for all tasks. They are as follows:

- Takeoff = 83.2%
- CAT II ILS Approach = 83.4%
- Missed Approach = 83.2%
- Enroute = 83.7%.

These data demonstrate that the evaluators could reliably detect the behavioral markers. Furthermore, the evaluators did not experience more difficulty identifying the performance markers for one task than for any of the others.

3.6.2 Subjective Ratings of Performance

In addition to indicating which performance markers they observed, the evaluators provided subjective ratings of overall performance on each maneuver for the crew as well as for each crew member. We analyzed these data to determine if the evaluators differed in how they evaluated overall performance and if they applied the rating scale reliably. Again, comparisons among all eight evaluators are difficult because they did not all observe the same crews. Therefore, inter-rater comparisons are limited to the two evaluators in each pair that observed the same crews.

In general, the evaluators perceived the crews and crew members to perform within standards. There were very few cases where an evaluator awarded an unsatisfactory rating; all of these were awarded for the same maneuver, the CAT II ILS Approach. One evaluator rated two crews (and both crew members) unsatisfactory; another evaluator rated one crew (and both crew members) unsatisfactory. Finally, a third evaluator rated a captain unsatisfactory on the same maneuver.

The evaluators varied in how strictly they applied the subjective rating scale. That is, some evaluators provided significantly lower ratings of proficiency than others. Table 3-4 depicts the average rating awarded by each evaluator for all crews, captains, and first officers. In addition, Table 3-4 presents the overall mean rating awarded by each evaluator and a test for differences between overall mean ratings provided by the two evaluators in each pair. In two pairs, the evaluators differed significantly from each other. That is, one evaluator in each of these two pairs perceived the flight crews as more proficient than the other evaluator in that pair.

Table 3-4

Average* Ratings of Proficiency

Evaluator	Crew*	Captain*	First Officer*	Overall**	t (46)
Pair 1: A	4.9	4.6	4.9	4.8	2.65
B	4.6	4.6	4.2	4.4	
Pair 2: A	5.1	4.8	5.4	5.1	1.40
B	4.6	4.5	4.7	4.6	
Pair 3: A	3.3	3.3	3.5	3.4	-3.34
B	4.1	3.9	4.3	4.1	
Pair 4: A	4.1	4.0	4.6	4.2	-.596
B	4.4	4.2	4.4	4.3	

* N=16 for each mean
**N=48 for each mean

Additionally, we evaluated the reliability between raters for assessing overall proficiency on individual tasks. Table 3-5 presents the values of Pearson correlation coefficients calculated on the ratings of overall proficiency awarded by the evaluators in each pair. All of the correlations are significant except one. The reliability between evaluators varied across pairs and crew members. The highest reliability was achieved for assessments of the captains; the ratings were least reliable when assessing the proficiency of first officers.

Table 3-5

Correlations Between Evaluator Proficiency Ratings for Each Pair of Evaluations (N=16).

Evaluator Pair	Crew	Captain	First Officer
1	.630**	.737**	.477
2	.570*	.633**	.542*
3	.692**	.591*	.540*
4	.668**	.772**	.499*
Mean	.640	.680	.510

* Denotes significant correlation where $P \leq .05$.
**Denotes significant correlation where $P \leq .01$.

These reliability measures appear somewhat low for a standardized cadre of instructor/evaluators. We expected to observe more standardization. Several factors may have contributed to these values. First, the evaluators received little training and no practice with the PMI prior to using it in a LOFT session. Clearly, more training and practice on the form and rating scale should improve reliability in future applications. Second, the evaluators clearly had more difficulty rating performance of the first officers. This may be due to the increased emphasis on evaluating the Captain during recurrent training. In addition, the evaluators experienced difficulty attributing crew errors to one team member or the other. Finally, the evaluators may have experienced difficulty because they do not normally assess varying degrees of proficiency for crews in recurrent training. The only formal discrimination they make is between very bad and acceptable. Thus, making relatively fine distinctions between different levels of good performance proved difficult.

3.6.3 Evaluator Opinions

The eight evaluators completed the opinion survey immediately after they observed their fourth crew perform the LOFT scenario. Reactions to the PMI were generally positive.

Table 3-6 presents the mean rating and standard deviation for each of the nine items with rating scales. A rating of one indicates that the evaluator strongly disagreed with the statement; a rating of five indicates that the evaluator strongly agreed with the statement. All of the mean ratings reflect positive opinions about the PMI and the ease with which the evaluators could rate the performance of individual crew members as well as intact crews. There was variability in the evaluators' opinions. However, very few provided extremely negative views. One evaluator indicated he strongly disagreed with two statements (4 and 9) and another evaluator indicated that he strongly disagreed with one statement (9).

Table 3-6

Mean Ratings and Standard Deviations for Survey Items 1-9

Survey Item	Mean	S.D.
1. Easy to use	3.2	1.1
2. Understand the markers	3.4	1.1
3. Easy to locate markers on form	3.6	0.9
4. Aided tracking performance	3.5	1.2
5. Aided performance assessment	3.5	0.8
6. Easy to rate individual pilots	3.5	0.8
7. Easy to rate crews	3.9	0.6
8. Significant training required	2.0	0.9
9. Well received by instructors	2.8	1.2

The evaluators generally wrote three types of responses to the open-ended questions. The first type suggested improvements to the PMI. These included a variety of comments about specific changes, particularly with reference to the performance markers. The comments included suggestions about adding or deleting specific markers, changing the wording, or putting them in a different sequence. All of the evaluators provided input about these kinds of refinements.

A second type of comment concerned the workload of the evaluator. Three of the evaluators commented that the PMIs placed an additional burden on the instructor/evaluator. All of the evaluators indicated they could manage the additional task, but this is an issue for future concern.

Finally, three of the evaluators indicated that they needed practice or training on how to use the PMI. They commented that their skill with the instrument improved with experience. In addition, they felt that we would need to provide structured training for instructors before incorporating this type of instrument into the recurrent curriculum.

4. CONCLUSIONS

This project had two specific objectives: to determine the utility of the CDM for identifying performance standards and to develop a prototype performance measurement instrument that integrates the measurement of CRM and technical flying skills. We have accomplished both of these objectives. This section discusses the conclusions we have drawn based on the data presented here as well as some observations that we made as we conducted the research. The conclusions are grouped under topical headings: the Critical Decision method, the Performance Measurement Instrument, the Analytic Process, Team Processes, and AQP.

4.1 Critical Decision Method

The CDM did not work well for eliciting information from the SMEs in this study. The effective use of CDM depends on recall of past incidents by experienced individuals. The richest incidents are those that were tough and challenged the individual's skills. These types of incidents provide opportunities to study the differences between experts and less- skilled performers.

In the present study, the interviewers were not able to elicit rich memorial incidents about the tasks of interest. SMEs related numerous stories about other types of incidents, but none directly related to missed approaches, ILS approaches, and engine failures. The SMEs did relate information about specific examples of good and poor performance that they had witnessed in the past. But, these generally referred to specific behaviors. They could not bring to the front the details of a rich, informative incident. Therefore, technically the interviews were not CDM interviews.

What did work well in this study was to focus the SMEs on their cognitive models of good and poor performance. We asked them to develop images of their perfect crew and how they would perform in this particular scenario. We asked the SMEs to describe the crew they wanted to emulate, the crew they wanted at the controls when their mother was a passenger. Similarly, we asked them to imagine the least proficient crew that worked for the airline. The SMEs generated mental simulations of these two crews performing the target scenario. These imagined crews probably represented a composite of various crews the SMEs had observed in the past. We then used the CDM probes to investigate these hypothetical incidents and were able to elicit data concerning the criteria that the SMEs use to distinguish good from poor performance.

Although the SMEs were not able to retrieve rich incidents from their own experience, we were able to accomplish our goals by modifying the technique. Focusing the SMEs on hypothetical situations allowed us access to some of the more subtle cues of performance to which the SMEs attend.

4.2 Performance Measurement Instrument

The PMI proved to be an effective and reliable means of recording and assessing both technical and team skills. The performance markers represent overt behaviors that reflect technical as well as cognitive and team skills. Instructors experienced little difficulty identifying and recording these markers. Furthermore, they were able to provide ratings of proficiency for individual crew members and the crew as a team. The instructors expressed positive attitudes about the utility and useability of the PMI. In future applications, additional practice and training about how to use the PMI should improve instructor utilization even more.

This approach to CRM integration and performance measurement meets the AQP requirements for assessing pilot proficiency and data collection. Performance measurement in AQP has at least two purposes. The first is to assess whether pilots have achieved the training objectives. The PMI provide this capability. It provides a further capability in that it allows assessment of crew proficiency. The second purpose of performance measurement is to assess the effectiveness of the training program. The data concerning pilot and crew proficiency will be analyzed periodically to assess whether the training program is effective.

However, simply knowing that the program is not effective is not enough. We must have diagnostic information as well to determine the source of the problem. Diagnostic information will allow program developers to identify which modules require revision and to make suggestions concerning the nature of those revisions. The PMI's performance markers provide this diagnostic capability. The performance markers must be selected carefully not only to represent the elements of expert assessment, but also to ensure that they provide sufficient information to diagnose training problems.

Finally, for the purposes of this study we chose to assess the performance markers using a dichotomous scale; the evaluator indicated whether he observed the behavior or not. This served the goal of determining if the evaluators understood the markers and could reliably identify them. We expect to expand these scales for some markers as we learn more about the nature of the behaviors.

4.3 Analytical Process

AQP has stimulated considerable interest in developing proficiency objectives for CRM skills and in integrating CRM and technical flying skills. Training developers have expended effort trying to fit the dimensions of CRM and its behavioral markers into their task analyses. These same people have experienced frustration at not accomplishing the task. There are several factors that contribute to this frustration: the nature of the behavioral markers, the types of skills required for good performance, the analytical methods employed, and the models of team performance that guide this work. We discuss each of these issues separately.

4.3.1 CRM Skills

CRM behavioral markers are neither specific nor homogenous. They are not specific in that each marker is not a specific behavior but represents a class of behaviors performed in a variety of tasks and conditions. The behaviors represented by a given marker vary depending on the task and existing set of conditions. The behavioral markers remain at one level of generality higher than specific behaviors. Thus, for a given marker, the training objectives will change from one task and condition set to another.

This is not consistent with developing effective training objectives. SPOs require specific performance statements and explicit statements of the standards for proficient performance for each subtask and condition set. Thus, analysts will need to identify specific indicators of CRM skills that are appropriate for given tasks and condition sets. They will identify the specific behaviors subsumed under each behavioral marker. These indicators can then be written into the training objectives.

CRM performance markers are not homogenous in that they represent different types of behavior. Some CRM skills comprise a set of basic interpersonal and communication skills that enable a group to function together as a team. That is, they overarch team performance (e.g., leadership-followership, assertion, command, tone, etc.). "Tone in the cockpit is friendly, relaxed, and supportive" serves as an example. This marker is of an interpersonal nature and includes a variety of behaviors that probably should occur throughout the flight, regardless of the task or conditions.

There is also a set of discrete team performance skills that can be identified for specific tasks under specific conditions (e.g., time horizon, time management, shared mental models, etc.). The behaviors represented by "Stays ahead of curve," vary from one situation to the next and may be very technical in nature. For example, in one case the crew may choose to remain at a lower airspeed because they anticipate being rushed on final approach. But, staying ahead of the curve in another situation may mean that the PF anticipates having to apply rudder when s/he changes the mode of flight. It may be more profitable to consider these different types of CRM skills separately. They may enter a task analysis at different levels and require different approaches to training.

4.3.2 Flying Skills

In conducting the front-end analyses for AQP, training developers have tended to lump CRM skills and cognitive skills (e.g., decision making, judgment, situation assessment, evaluations, perceptions) together. This has contributed to the problem of analyzing these skills for the purpose of developing training objectives. Cognitive skills are not CRM skills nor does CRM training provide students with cognitive skills. As Foushee and Helmreich (1988) state, CRM training focuses on the input factors of team performance, not the process factors. CRM training is important and necessary, but it is designed to produce an environment that will facilitate cognition of the individual and the team. It is not designed to affect the cognitive processes themselves. To develop training that will enhance cognitive processes, we must address these issues directly.

Flying has always involved cognitive processing. However, the recent increase in the use of computers and automation has changed and complicated the cognitive skills required to fly modern aircraft. Numerous popular studies have brought attention to the importance of cognitive skills necessary for flying and the impact that automated aircraft have had on pilot performance (e.g., Wiener, 1989; Wiener, et al., 1991). Other researchers have investigated the cognitive demands placed on the flight crew by a single piece of equipment, the flight management system (e.g., Kaempf, Klein, & Thordsen, 1991; Sartor & Woods, 1991). These studies all indicate that the cognitive task requirements have a significant impact on crew performance and on aviation safety. The advanced technologies create problems for pilots who learned to fly on analogue airplanes and affect how student pilots should learn how to fly. A significant contribution that AQP can make to the aviation industry is to provide a means for developing training objectives for cognitive tasks and for developing training programs specifically designed to enhance cognitive performance.

4.3.3 Analytic Method

How can we train flight crews to make better decisions? How can training affect the mental models of flying and the mental models they maintain about what the automation is doing for them? How can training ensure that crew members share the same accurate models?

Classical behavioral task analyses do not provide the answers to these questions. BTA focuses only on the overt behaviors exhibited by the crew. Some analysts have begun to include such verbs as "Decide" and "Assess" at the lowest level of their analytical hierarchy. However, this does not constitute an analysis of cognitive processes, nor does it provide sufficient data to support training objectives. An analytical method designed specifically for identifying the cognitive demands of tasks is needed.

Cognitive Task Analyses (CTAs) serve this purpose. CTAs evolved from the instructional design process. Researchers recognized that BTAs adequately described jobs that comprise tasks largely psychomotor in nature and with fixed procedural sequences. However, increasingly sophisticated equipment and complex environments have created tasks with large decision-making components that place high cognitive demands on the individual. BTAs do not provide adequate models of these types of jobs. Thus, training developers needed methods to describe the cognitive components of jobs as well as the behavioral.

Numerous researchers have reported success with CTA for developing training programs (e.g., Redding, 1989; Ryder, Redding, & Beckshi, 1987) and for developing system interfaces (e.g., Kaempf, Wolf, Thordsen, & Klein, 1992; Roth & Woods, 1990; Woods & Hollnagel, 1987). BTAs are directed at the objective task performance, whereas CTAs are directed at the psychological processes underlying the performance. CTAs provide a means of understanding the cognitive demands of a task, for identifying desired cognitive performance, and for developing training objectives designed to effect that desired cognitive performance.

Cognitive task analyses should be integrated into the AQP process to provide a complete picture of pilot and crew task requirements. This will enable training developers to produce training objectives that accurately address cognitive performance and to develop training that will shape the cognitive processes underlying overt performance.

4.4 Team Processes

Finally, part of the difficulty experienced in the AQP process has been due to the absence of an adequate model of team performance. Again, CRM training focuses on the input factors of team performance. CRM training indirectly affects team output by creating an environment that facilitates team performance. However, little is known about the processes or mechanisms of team performance. Therefore, we do not yet have a theoretical base for developing training objectives that attempt to directly manipulate team processes.

Zsambok, Klein, Kyne, and Klinger (1992) have recently developed a model that describes how teams make decisions, that describes the ontogeny of team decision-making skills, and that provides a framework for effecting positive changes in team performance. This model may also serve as a central model of team processes for the flight crew.

The model of team decision making describes the team as an entity that develops through training and experience. The team develops along three dimensions: team identity, team conceptual level, and team self-monitoring. Team identity refers to the extent to which members conceive of the team as an interdependent unit and operate from that perspective while engaged in their tasks. Team conceptual level describes how the team thinks, solves problems, makes decisions, and takes actions collectively. A key concept of this dimension is the "team mind:" teams engage in cognitive activities analogous to those of the individual. Team self-monitoring describes the regulatory process that helps progress on the other two dimensions. Teams can advance from lax to vigilant self-monitoring, from weak to strong team identity, and from a low to high conceptual level.

Zsambok et al. (1992) have identified observable markers that reflect a team's level of performance on each of the three dimensions. These team behaviors are key elements of the team decision-making process and can be used to distinguish good decision-making teams from poor ones. The markers include cognitive processes essential for successful outcomes including time horizon, time management, shared situation assessments (mental models), and defining roles and functions. In the present study, we observed how the SMEs considered these cognitive skills as essential in evaluating crew and pilot performance. The good crews anticipate and plan for upcoming events and they ensure that both crew members have the same mental model. Many of the CRM-related behaviors serve only to facilitate these processes; they fail to provide a means of describing them.

Zsambok et al. have demonstrated the reliability and utility of these markers. Furthermore, they have demonstrated the effectiveness of this model for developing training programs designed to enhance team performance.

The model of advanced team decision making may provide a framework for understanding the team processes of a flight crew. By understanding these processes, we will be able to develop more focused and effective methods for enhancing team skills.

4.5 AQP

AQP provides an opportunity for air carriers to revitalize their training programs by incorporating advances in training technology and cognitive science and by tailoring their training programs to suit their specific needs. This will require creative solutions and may require significant changes in the way the airlines do business. For example, many airlines have separated their ground-training and flight-training programs into separate divisions that compete for resources. Successful implementation of AQP will require an integration of the two that focuses on the goal of producing the best crews for the least expense.

AQP also provides opportunities to study and resolve issues concerning team and pilot training. Several issues have come to the fore, those concerning CRM integration and cognitive task requirements. Development of AQP training in the aviation industry provides us an excellent opportunity for enhancing and implementing these technologies.

5. REFERENCES

Anderson, J. R. (1982). "Acquisition of cognitive skill." Psychological Review, 89(4), 369-406.

Berry, D. C. (1987). "The Problem of Implicit Knowledge." Expert Systems, 4(3), 144-151.

Brezovic, C. P., Klein, G. A., & Thordsen, M. (1987). Decision Making in Armored Platoon Command (KATR-858(b)-87-05F). Yellow Springs, OH: Klein Associates Inc. Prepared under contract MDA903-85-C-0327 for U.S. Army Research Institute, Alexandria, VA.

Calderwood, R., Crandall, B. W., & Klein, G. A. (1987). Expert and Novice Fireground Command Decisions (ADB154-224). Alexandria, VA: Defense Logistics Agency.

Chi, M. T. H., Glaser, R., & Farr, M. J. (1988). The Nature of Expertise. Hillsdale, NJ: Lawrence Erlbaum Associates.

Chidester, T. R., Helmreich, R. L., Gregorich, S. E., & Geis, C. E. (1991). "Pilot Personality and Crew Coordination: Implications for Training and Selection." The International Journal of Aviation Psychology, 1,(1), 25-44. Hillsdale, NJ: Lawrence Erlbaum Associates.

Childs, J. M., Spears, W. D., & Prophet, W. W. (1983). Private Pilot Flight Skill Retention, 6, 8, and 24 Months Following Certification (DOT/FAA-CT-83-34). Washington, DC: Federal Aviation Administration.

Cooper, G. E., White, M. D., & Lauber, J. K. (Eds.). (1979). Resource Management on the Flight Deck: Proceedings of a NASA/Industry Workshop. Moffett Field, CA: National Aeronautics and Space Administration.

Crandall, B. (1989). "A Comparative Study of Think-aloud and Critical Decision Knowledge Elicitation." ACM SIGART, 108, April.

Crandall, B., & Calderwood, R. (1989). Clinical Assessment Skills of Experienced Neonatal Intensive Care Nurses. Yellow Springs, OH: Klein Associates Inc. Final Report prepared for the National Center for Nursing Research, National Institutes for Health under Contract No. 1 R43 NR01911 01.

Crandall, B., & Klein, G. A. (1988). Key Components of MIS Performance. Report to Standard Register. Yellow Springs, OH: Klein Associates Inc.

Deutsch, R. D. (1990). "War Stories: A Knowledge Base for Aiding Command and Control Decision-Making in the New Global Context." Unpublished manuscript submitted to Office of Naval Research, Cognitive and Neural Sciences Division, December.

Federal Aviation Administration. (1991a). Advanced Qualification Program. Special Federal Aviation Regulation 58. Washington, DC: Federal Aviation Administration.

Federal Aviation Administration. (1991b). Advanced Qualification Program. Advisory Circular 120-54. Washington, DC: Federal Aviation Administration.

Federal Aviation Administration. (1992). Crew Resource Management (CRM) Advisory Circular (AC 120-51A). Washington, DC: Federal Aviation Administration.

Fiedler, F. E. (1967). A Theory of Leadership Effectiveness. NY: McGraw-Hill.

Fitts, P. M., & Posner, M. I. (1967). Human Performance. Belmont, CA: Brooks Cole.

Flanagan, J. C. (1954). "The Critical Incident Technique." Psychological Bulletin, 51, 327-358.

Foushee, H. C. (1984). "Dyads and Triads at 35,000 Feet: Factors Affecting Group Process and Aircrew Performance." American Psychologist, 39, 885-893.

Foushee, H. C. (1985). "Realistic Training for Effective Crew Performance." In Proceedings, 4th Aerospace Behavioral Engineering Technology Conference. Warrendale, PA: SAE, Inc.

Foushee, H. C., & Helmreich, R. L. (1988). "Group Interaction and Flight Crew Performance." In E. L. Wiener and D. C. Nagel (Eds.), Human Factors In Aviation (189-227). NY: Academic Press, Inc.

Foushee, H. C., & Manos, K. L. (1981). "Information Transfer Within the Cockpit: Problems in Intracockpit Communications." In C. E. Billings and E. S. Cheaney (Eds.), Information Transfer Problems in the Aviation System (NASA Technical Paper 1875). Moffett Field, CA: NASA Ames Research Center.

Greer, G. D., Smith, J. F., & Hatfield, J. L. (1962). Improving Flight Proficiency Evaluation in Army Helicopter Pilot Training (HumRRO Technical Report 69-26). Alexandria, VA: Human Resources Research Organization.

Helmreich, R. L. (1982, August). "Pilot Selection and Training." Paper presented at the meeting of the American Psychological Association, Washington, DC.

Helmreich, R. L. (1984). "Cockpit Management Attitudes." Human Factors, 26, 583-589.

Helmreich, R. L. (1986). "Theory Underlying CRM Training: Psychological Issues in Flight Crew Performance and Crew Coordination." In H. W. Orlady and H. C. Foushee (Eds.), Cockpit Resource Management Training: Proceedings of a NASA/MAC Workshop (NASA CP 2455), 15-22. Moffett Field, CA: NASA Ames Research Center.

Helmreich, R. L. (1987). "Theory Underlying CRM Training: Psychological Issues in Flightcrew Performance and Crew Coordination." In H. W. Orlady and H. C. Foushee (Eds.), Cockpit Resource Management Training: Proceedings of the NASA/MAC Workshop. Moffett Field, CA: NASA-Ames Research Center.

Helmreich, R. L., Foushee, H. C., Benson, R., & Russini, R. (1986). "Cockpit Management Attitudes: Exploring the Attitude-Performance Linkage." Aviation, Space, and Environmental Medicine, 57, 1198-1200.

Helmreich, R. L., Hackman, J. R., & Foushee, H. C. (1988). Evaluating Flight Crew Performance: Policy, Pressures, Pitfalls, and Promise (NASA Technical Memorandum). Moffett Field, CA: NASA Ames Research Center.

Helmreich, R. L., Wilhelm, J. A., & Gregorich, S. E. (1988). Revised Versions of the Cockpit Management Attitudes Questionnaire (CMAQ) and CRM Seminar Evaluation Form (NASA/UT Technical Report 88-3). Austin, TX: University of Texas at Austin.

Helmreich, R. L., Wilhelm, J. A., Gregorich, S. E., & Chidester, M. A. (1990). "Preliminary Results from the Evaluation of Cockpit Resource Management Training: Performance Ratings of Flightcrews." Aviation, Space, and Environmental Medicine, 61, 576-579.

Helmreich, R. A., Wilhelm, J. A., Kello, J. E., & Taggant, W. R. (1987). Reinforcing and Evaluating Crew Resource Management: Check Airman/LOFT Instructor Reference Manual (NASA-UT Technical Manual 90-1). Austin, TX: The University of Texas at Austin.

Hoffman, R. R. (1987). "The Problem of Extracting the Knowledge of Experts from the Perspective of Experimental Psychology." AI Magazine, 8, 53-67.

Howell, W. C., & Cooke, N. J. (1989). "Training the Human Information Processor: A Look at Cognitive Models." In I. L. Goldstein (Ed.), Training and Development in Work Organizations: Frontiers of Industrial and Organizational Psychology, 121-182. San Francisco, CA: Jossey-Bass.

Hutchins, E. (1991). "How a Cockpit Remembers Its Speed." In Proceedings of Second Ames AS/A Program Investigator's Meeting. Moffett Field, CA: NASA Ames Research Center.

Kaempf, G. L., & Blackwell, N. J. (1990). Transfer of Training Study of Emergency Touchdown Maneuvers in the AH-1 Flight and Weapons Simulator (Research Report 1561). Alexandria, VA: Army Research Institute for the Behavioral and Social Sciences.

Kaempf, G. L., Cross, K. D., & Blackwell, N. J. (1989). Backward Transfer and Skill Acquisition in the AH-1 Flight and Weapons Simulator (Research Report 1537). Alexandria, VA: Army Research Institute for the Behavioral and Social Sciences.

Kaempf, G. L., Klein, G. A., & Thordsen, M. L. (1991). Applying Recognition-primed Decision-making to Man-Machine Interface Design (Final Technical Report). Fairborn, OH: Klein Associates Inc. Prepared under contract NAS2-13359 for NASA-Ames Research Center, Moffett Field, CA.

Kaempf, G. L., Wolf, S., Thordsen, M. L., & Klein, G. (1992). Decision-making in the AEGIS Combat Information Center. Fairborn, OH: Klein Associates Inc. Prepared under contract N66001-90-C-6023 for the Naval Command, Control and Ocean Surveillance Center, San Diego, CA.

Kanki, B. G., Lozito, S., & Foushee, H. C. (1987). "Communication Indexes of Crew Coordination." In R. S. Jensen (Ed.), Proceedings of the Fourth International Symposium on Aviation Psychology, 406-412. Columbus, OH: Association of Aviation Psychologists.

Klein, G. A., Calderwood, R., & MacGregor, D. (1989). "Critical Decision Method for Eliciting Knowledge." IEEE Transactions on Systems, Man, and Cybernetics, 19(3), 462-472.

Krumm, R. L., & Farina, A. J. (1962). Effectiveness of Integrated Flight Simulator Training in Promoting B-52 Crew Coordination (MRL-TDR-62-1). Wright-Patterson AFB, OH: Air Force Systems Command.

Lofaro, R. J. (1992). Workshop on Integrated Crew Resource Management (CRM) (DOT/FAA/RD-92/5). Washington, DC: Federal Aviation Administration.

Norman, D. A. (1991). "Cognitive Science in the Cockpit." CSERIAC Gateway, 2(2), 1-6.

Orasanu, J. M. (1990). Shared Mental Models and Crew Decision-making (CSL Report 46). East Rutherford, NJ: Princeton University.

Povenmire, H. K. (1989). Cockpit Resource Management: A Selected Annotated Bibliography (AFHRL-TP-89-22)." Brooks Air Force Base, TX: U.S. Air Force Human Resources Laboratory.

Povenmire, H. K., Rockway, M. R., Bunecke, J. L., & Patton, M. W. (1989). "Cockpit Resource Management Skills Enhance Combat Mission Performance in a B-52 Simulator." In R. S. Jensen (Ed.), Proceedings of the Fifth International Symposium on Aviation Psychology, 489-494. Columbus, OH: Association of Aviation Psychologists.

Prophet, W. W., & Jolley, O. B. (1969). Evaluation of the Integrated Contact-Instrument Concept for Army Fixed Wind Flight Instruction (HumRRO Technical Report 69-26). Alexandria, VA: Human Resources Research Organization.

Redding, R. E. (1989). Perspectives on Cognitive Task Analysis: The State of the Art. Proceedings of the Human Factors Society 33rd Annual Meeting, 1348-1352.

Redding, R. E. (1990). "Taking Cognitive Task Analysis into the Field Bridging the Gap from Research to Application." Proceedings of the Human Factors Society 34th Annual Meeting, 2, 1304-1308. Santa Monica, CA: Human Factors Society.

Redding, R. E., & Lierman, B. (1990). "Development of a Part-Task, CBI Trainer Based Upon a Cognitive Task Analysis." Proceedings of the Human Factors Society 34th Annual Meeting, 1337-1341.

Roth, E. M., & Woods, D. D. (1990). "Analyzing the Cognitive Demands of Problem-Solving Environments: An Approach to Cognitive Task Analysis." Proceedings of the Human Factors Society 34th Annual Meeting, 2, 1314-1317.

Ruffell Smith, H. P. (1979). "A Simulator Study of the Interaction of Pilot Workload with Errors, Vigilance, and Decisions (Technical Memorandum 78482)." Moffett Field, CA: NASA-Ames Research Center.

Ryder, J. M., Redding, R. E., & Beckshi, P. F. (1987). "Training Development for Complex Cognitive Tasks." Proceedings of the Human Factors Society 31st Annual Meeting, 1261-1265.

Sams, T. L. (1987). "Cockpit Resource Management Concepts and Training Strategies." Unpublished doctoral dissertation, East Texas State University, TX.

Sartor, N. B., & Woods, D. D. (1991). "Situation Awareness: A Critical But Ill-defined Phenomenon." The International Journal of Aviation Psychology, 1(1), 45-57.

Schank, R. C. (1990). Tell Me a Story. NY: Charles Scribner's & Sons.

Schlager, M. S., Means, B., & Roth, C. (1990). "Cognitive Task Analysis for the Real(Time) World." Proceedings of the Human Factors Society 34th Annual Meeting, 1309-1313.

Schneider, W., Vidulich, M., & Yeh, Y. (1982). "Training Spatial Skills for Air Traffic Control." Proceedings of the Human Factors Society 26th Annual Meeting, 10-14.

Shelnutt, J. J., Spears, W. D., & Prophet, W. W. (1981). Effects of Different Distributions of Training Time on the Acquisition of Private Pilot Flying Skills (FAA-CT-81-76). Washington, DC: Federal Aviation Administration.

Smith, J. F., Flexman, R. E., & Houston, R. C. (1952). Development of an Objective Method of Recording Flight Performance (Technical Report 52-15). Lackland Air Force Base, TX: Air Training Command, Human Resources Research Center.

Smith, K. M. (1991). "CRM Integration." Paper presented at the meeting of the Human Factors Steering Committee, Denver, CO.

Taynor, J., Crandall, B., & Wiggins, S. (1987). The Reliability of the Critical Decision Method (KATR-863(B)-87-07F). Yellow Springs, OH: Klein Associates Inc. Prepared under contract MDA903-86-C-0170 for the U.S. Army Research Institute Field Unit, Alexandria, VA.

Thordsen, M. L., & Calderwood, R. (1989). Protocol Analysis of Flight Crew Decision-Making During Simulated Malfunctions. Yellow Springs, OH: Klein Associates Inc. Prepared under Order No. A72145C for the NASA/Ames Research Center, Moffett Field, CA.

Thordsen, M. L., Klein, G. A., & Wolf, S. (1990). Observing Team Coordination Within Army Rotary-wing Aircraft Crews (Final Technical Report). Yellow Springs, OH: Klein Associates Inc. Prepared under contract MDA903-87-C-0523 for the U.S. Army Research Institute, Aviation Research and Development Activity, Ft. Rucker, AL.

Thornton, R. C., Kaempf, G. L., Zeller, J. L., & McAnulty, D. M. (1991). An Evaluation of Crew Coordination and Performance During a Simulated UH-60 Helicopter Mission (ASI690-344-91B). Fort Rucker, AL: Anacapa Sciences, Inc.

Wheale, J. L. (1983). "Crew Coordination and Interaction on the Flight Deck." Proceedings of the 31st International Congress of Aviation and Space Medicine, 88-102. Dublin: Aer Lingus.

Wiener, E. L. (1989). Human Factors of Advanced Technology ("Glass Cockpit") Transport Aircraft (NASA Contractor Report 177528). Moffett Field, CA: Ames Research Center.

Wiener, E. L., Chidester, T. R., Kanki, B. G., Palmer, E. A., Curry, R. E., & Gregorich, S. E. (1991). The Impact of Automation on Crew Coordination and Communication: Overview, LOFT Evaluations, Error Severity, and Questionnaire Data. NASA Contractor Report (177587). Moffett Field, CA: National Aeronautics and Space Administration.

Woods, D. D., & Hollnagel, E. (1987). "Mapping Cognitive Demands in Complex Problem-Solving Worlds." International Journal of Man-Machine Studies, 26, 257-275.

Zsambok, C. E., Klein, G., Kyne, M. M., & Klinger, D. W. (1992). Advanced Team Decision-making: A Developmental Model. Fairborn, OH: Klein Associates Inc. Prepared by Klein Associates for the U.S. Army Research Institute for the Behavioral and Social Sciences, Contract MDA903-90-C-0117.

APPENDIX A

DIMENSIONS OF CREW RESOURCE MANAGEMENT (CRM) SKILLS

Briefing (Conduct and Quality)

Description

The effective briefing will be operationally thorough, interesting, and will address coordination, planning, and problems. [Although primarily a Captain's responsibility, other crew members may add significantly to planning and definition of potential problem areas.]

Behavioral Markers

1. Establishes environment for open/interactive communications (e.g., calls for questions or comments, answers questions directly, listens with patience, does not interrupt or "talk over," does not rush through the briefing, makes eye contact as appropriate).

2. Is interactive, two-way, and emphasizes the importance of questions, critique, and the offering of information.

3. Establishes "team concept" (e.g., uses "we" language, encourages all to participate and help with the flight).

4. Covers pertinent safety and operational issues.

5. Identifies potential problems such as weather, delays, and abnormal system operations.

6. Provides guidelines for crew actions--division of labor and crew workload addressed.

7. Includes cabin crew as part of team in the briefing, as appropriate.

8. Sets expectations for how deviations from SOP are to be handled.

9. Establishes policy guidelines for the operation of automated system (i.e., when system will be disabled, programming actions that must be verbalized and acknowledged).

10. Specifies PF and PNF duties and responsibilities with regard to Flight Management System.

Inquiry/Advocacy/Assertion (Practiced)

Description

This rating assesses the extent to which crew members advocate the course of action they feel best, even when it involves conflict and disagreements with others.

Behavioral Markers

1. Crew members speak up, and state their information with appropriate persistence, until there is some clear resolution and decision.

2. "Challenge and response" environment is developed.

3. Questions are encouraged, and are answered openly and nondefensively.

4. Crew members are encouraged to ask questions regarding crew actions and decisions.

5. Crew members seek information and direction from others when necessary.

6. Crew members question status and programming of Flight Management System to verify situational awareness.

Crew Self-Critique (Decisions and Actions)

Description

This item evaluates the extent to which crew members conduct and participate in a debriefing, operation review, and critique of activities, which includes the product, the process, and the people involved. Critique can, and should, occur during an activity, and/or after completion of the activity.

Behavioral Markers

1. Given at appropriate times, both low and high workload.
2. Deals with positive as well as negative aspects of crew performance during flight.
3. Interactively involves the whole crew.
4. Made a positive learning experience--feedback is specific, objective, based on observable behavior, and given constructively.
5. Accepted objectively and nondefensively.

Communications/Decisions

Description

This rating reflects the extent to which free and open communication is practiced. It includes the extent to which crew members provide necessary information at the appropriate time (for example, initiating checklists, alerting others to developing problems). Active participation in decision-making process encouraged and practiced. Questioning of actions and decisions is proper. Decisions made are clearly communicated and acknowledged.

Behavioral Markers

1. Operational decisions are clearly stated to other crew members.

2. Crew members acknowledge understanding of decisions made.

3. "Bottom lines" are established and communicated for safety of operations.

4. The "big picture" and the game plan are shared within the team including Flight Attendants and others.

5. Crew members are encouraged to state their own ideas, opinions, and recommendations.

6. Effort is made to provide an atmosphere conducive to open and free communications.

7. Entries and changes to Flight Management System parameters are verbalized and acknowledged.

Leadership-Followership/Concern for Tasks

Description

This rating evaluates the extent to which appropriate leadership and followership is practiced. It reflects the extent to which the crew is concerned with the effective accomplishment of necessary tasks.

Behavioral Markers

1. Utilize all available resources to accomplish job at hand.

2. Coordinate flight deck activities to establish proper balance between authority and assertiveness.

3. Act decisively when the situation requires.

4. Demonstrate desire to achieve most effective possible operation.

5. Recognize need to maintain adherence to SOPs.

6. Ensure that group climate is appropriate to operational situation (i.e., social conversation in low workload conditions but not high).

7. Recognize effects of stress and fatigue on performance.

8. Manage time available for task accomplishment.

9. Recognize and deal with demands on resources posed by operation of Flight Management System.

10. Disengage Flight Management System operation when programming demands could reduce situational awareness or create work overloads.

Interpersonal Relationships/Group Climate

Description

This evaluation reflects the quality of observed interpersonal relationships among and the overall climate of the flight deck. This is independent of demonstrated concern with accomplishment of required tasks.

Behavioral Markers

1. Remain calm under stressful conditions.

2. Show sensitivity and ability to adapt to other crew members' personalities and personal characteristics.

3. Recognize symptoms of psychological stress and fatigue in self and others (e.g., note when a crew member is not communicating, and draw him/her back into the team; recognize when they are experiencing "tunnel vision," and seek help from the team).

4. "Tone" in the cockpit is friendly, relaxed, supportive.

5. During times of low communication, crew members check in with each other to see how they are doing.

Preparation/Planning/Vigilance

Description

This rating indicates the extent to which crews anticipate contingencies and actions that may be required. Excellent crews are always "ahead of the curve" while poor crews continually play catch-up. Vigilant crews devote appropriate attention to required tasks and respond immediately to new information. A crew indulging in casual social conversation during periods of low workload is not necessarily lacking in vigilance if flight duties are being discharged properly.

Behavioral Markers

1. Demonstrate and express situational awareness--the "model" of what is happening is shared within the crew.

2. Monitoring of all instruments and communications, sharing relevant information with the rest of the crew.

3. Monitor climatic and traffic conditions, sharing relevant information with the rest of the crew.

4. Avoid "tunnel vision" of stress--stating or asking for the "big picture."

5. Be aware of factors such as stress that can reduce vigilance--thus, monitoring the performance of other crew members.

6. Stay "ahead of curve" in preparing for expected or contingency situations (including approaches, weather, etc.).

7. Verbally insure that cockpit and cabin crew are aware of plans.

8. Include all appropriate crew members in planning process.

9. Plan for sufficient time prior to maneuvers for programming of Flight Management Computer.

10. Ensure that all crew members are aware of status and changes in Flight Management System parameters.

Workload Distributed/Distractions Avoided

Description

This is a rating of time and workload management. It reflects how well the crew managed to distribute the tasks and avoid overloading individuals. It also considers the ability of the crew to avoid being distracted from essential activities and how work is prioritized.

Behavioral Markers

1. Crew members report and admit work overloads.

2. Take action to distribute tasks to maximize efficiency.

3. Clearly communicate workload distribution and acknowledge.

4. Make sure that non-operational factors such as social interaction do not interfere with necessary task duties.

5. Communicate the work priorities clearly to the crew.

6. Make sure that secondary operational tasks (i.e., dealing with passenger needs, company communications) are prioritized so as to allow sufficient resources for dealing effectively with primary flight duties.

7. Recognize and report overloads in others.

8. Crew members recognize potential distractions posed by Flight Management Systems and take appropriate preventive action, including disengaging.

APPENDIX B

SURVEY OF EVALUATOR OPINIONS

Survey of Evaluator Opinions

NOTE:

This survey contains 14 items for you to complete. All of the items refer to the Flight Crew Evaluation Form that you recently used as you conducted recurrent training. Please take a few minutes to consider your answers carefully and provide as much accurate feedback as you can. Items 1-9 ask you to provide a rating of how much you agree with the given statement. Items 10-14 are open-ended questions and ask you to provide more elaborate feedback about the form. Thank you for your cooperation and help.

Your Name: ____________________

1. I found the form easy to use when in the simulator.

Strongly Disagree				Strongly Agree
1	2	3	4	5

2. I was able to quickly understand the performance markers.

Strongly Disagree				Strongly Agree
1	2	3	4	5

3. I found that it was easy for me to locate the markers that I needed on the form.

Strongly Disagree				Strongly Agree
1	2	3	4	5

4. I found the form helped me to keep track of how the crew performed during the session.

Strongly Disagree				Strongly Agree
1	2	3	4	5

5. I believe that the form helped me to make a more accurate determination of crew performance.

Strongly Disagree				Strongly Agree
1	2	3	4	5

6. I found that it was easy to determine performance ratings for the individual crew members.

Strongly Disagree				Strongly Agree
1	2	3	4	5

7. I found that it was easy to determine performance ratings for the crew as a team.

Strongly Disagree				Strongly Agree
1	2	3	4	5

8. Training instructors to use this form would require a significant amount of training.

Strongly Disagree				Strongly Agree
1	2	3	4	5

9. I believe that this form would be well received by instructor pilots.

Strongly Disagree				Strongly Agree
1	2	3	4	5

10. What behaviors would you like to see added to the list of markers?

11. Are there any markers on the list that should not be there? If so, which ones?

12. Are there any markers that simply don't make sense to you? If so, which ones?

13. Are there any characteristics of this form which make it hard for you to use in the simulator?

14. Do you believe this form would aid you or impede you while conducting training in the simulator? Please explain.

APPENDIX C

TERMINAL AND SUPPORTING PROFICIENCY OBJECTIVES

Terminal Proficiency Objective

Task: Preflight/Before Start Operations

Performance: Perform the task Preflight/Before Start Operations including five subtasks: Initial Power Up Procedures, First Flight of the Day Test Procedures, Exterior Preflight Procedures, Interior Preflight Procedures, Departure Briefing.

Conditions: To be performed in a B-757/767 aircraft or flight simulator under any conditions that exist after the environment and aircraft were considered acceptable during the Preflight Planning Procedures.

Standards: Existing conditions do not significantly alter the performance or standards for the specific subtasks. Therefore, one SPO exists for each of the subtasks subsumed under this task. Flight crews should accomplish the performance standards for each of these SPOs.

Supporting Proficiency Objective

Subtask: Departure Briefing

Performance: The PF conducts the briefing and includes the following information: takeoff minimums, takeoff alternate requirements, low visibility procedures, noise abatement procedures, rejected takeoff plan, runway and special terrain conditions, and SID/special Jeppesen pages.

Conditions: To be performed in a B-757/767 aircraft or flight simulator under any conditions that exist after the environment and aircraft were considered acceptable during the Preflight Planning Procedures. The other subtasks of the task Preflight/Before Start Operations have been completed.

Standards: In addition to the required elements of the briefing, the PF should include any unusual or abnormal items that may reasonably be anticipated. The flight crew should use the briefing as an opportunity to establish command leadership roles with the crew and to foster team relationships. The crew should not rush the briefing and should ensure that all points are understood by both pilots. The briefing has two active participants. The PNF should ask questions and elaborate to ensure that both pilots share a common understanding.

Terminal Proficiency Objective

Task: Takeoff

Performance: Perform Takeoff Task to include four subtasks: Assess Performance and Environmental Factors, Takeoff Roll Procedures, Rotation and Liftoff Procedures, and Initial Climb Procedures.

Conditions: The Takeoff Task may be performed under a number of environmental and aircraft conditions to include normal weather, crosswinds during Takeoff Roll and Rotation and Liftoff, wind shear during any subtask of the Takeoff Task, single-engine failure at any point in the Takeoff Task, and flap and landing gear malfunctions at any point in the Takeoff Task.

Standards: Performance standards for the subtasks Takeoff Roll Procedures, Rotation and Liftoff Procedures, and Initial Climb Procedures are determined by the existing conditions. The existing conditions alter significantly the performance or standards for the subtasks. Therefore, separate supporting proficiency objectives exist for each subtask that is affected by each set of conditions. Flight crews should be able to meet the performance standards specified in each of these supporting proficiency objectives.

Supporting Proficiency Objective

Subtask: Assess Performance and Environmental Factors

Performance: The flight crew verifies that no changes have been made to the takeoff plan. The flight crew evaluates the aircraft gross weight, runway length and conditions, and the environmental conditions to ensure that the aircraft is configured properly for the conditions and that the appropriate procedures will be used during the takeoff. The PF completes the Takeoff Briefing including assigned heading, airspeed, altitude, and any unusual conditions that were not briefed previously.

Conditions: To be performed in a B-757/767 aircraft or flight simulator with no aircraft systems malfunctions, the aircraft positioned at the end of the runway, the Taxi Checklist complete, and the Before Takeoff Checklist partially complete. Any environmental conditions may exist.

Standards: The flight crew should assess all relevant factors paying particular attention to those that may have developed during taxi. The flight crew should ensure that the aircraft is configured properly for the conditions. The PF should brief the PNF on all points required for the Takeoff Briefing.

Supporting Proficiency Objective

Subtask: Takeoff Roll Procedures

Performance: The PNF selects Heading Hold on the MCP. The PF smoothly applies equal power to both engines toward takeoff power. Throughout the takeoff roll, the PF should monitor the aircraft ground track and ensure that the aircraft remains close to the runway center line by applying appropriate pressure to the rudder and ailerons while maintaining slight forward pressure on the yoke. As the N_1 EPR reaches 1.1, the PF calls for and the PNF selects EPR on the MCP. The Captain must remain alert for any conditions that warrant rejecting the takeoff. To facilitate rejecting the takeoff, the Captain uses the right hand to maintain positive control of the throttle until the aircraft accelerates through V_1. Performance standards for rejecting the takeoff are identified under the SPO for the subtask Rejected Takeoff. At 80 KIAS, the PNF notifies the PF of the airspeed, the engine instruments' conditions, and the Throttle Hold is active on the EICAS. As the aircraft accelerates, the PNF notifies the PF when passing through V_1, V_r, and V_2

Conditions: To be performed in a B-757/767 aircraft or flight simulator with no aircraft systems malfunctions, Departure Briefing complete, Taxi and Before Takeoff Checklists complete, the aircraft aligned on the runway center line, and cleared for takeoff by ATC. Crosswinds may exist on the runway; no other abnormal weather conditions exist.

Standards: During the takeoff roll, aircraft alignment should not deviate more than 20 feet to either side of the runway center line. The PF should monitor the flight instruments and the environment outside the aircraft. The PNF should monitor the flight and systems' instruments and support the PF as necessary. The flight crew should accomplish all required callouts and acknowledgements.

Supporting Proficiency Objective

Subtask: Rotation and Liftoff Procedures

Performance: The PF maintains runway alignment and the PNF monitors the flight instruments and the status of aircraft systems. The PNF notifies the PF when accelerating through V_r and V_2. At V_r, the PF smoothly applies aft pressure on the yoke to rotate the aircraft at approximately 3° per second toward a target pitch attitude of 15°. After V_2, the PNF monitors the barometric altimeter and notifies the PF when the altimeter indicates a positive rate of climb. The PF then calls for and the PNF retracts the landing gear.

Conditions: To be performed in a B-757/767 aircraft or flight simulator with no aircraft systems malfunctions, aircraft aligned on the runway center line, and the aircraft accelerating through V_r. Crosswinds may exist on the runway; no other abnormal weather conditions exist.

Standards: The PF should maintain assigned heading ± 5°. The PF should smoothly rotate the aircraft to a pitch attitude not to exceed 18° and adjust the pitch attitude to achieve a target minimum airspeed of V_2 + 15 KIAS. The PF should monitor the flight instruments and the outside environment; the PNF should monitor the flight and systems' instruments and support the PF as necessary. The flight crew should accomplish all required callouts and acknowledgements.

Supporting Proficiency Objective

Subtask: Initial Climb Procedures

Performance: The PF maintains the assigned heading and a target minimum airspeed of V_2 + 15. The PNF monitors the aircraft systems and flight instruments and supports the PF as necessary. As the aircraft approaches 1000 feet AGL, the PF decreases the pitch attitude to approximately 10 degrees and directs the PNF to select Flight Level Change on the MCP, to set V_{ref} 30 + 80 on the Command Airspeed Bug, and to select climb power. The PF calls for and the PNF retracts the flaps on speed schedule. Once the flaps have been retracted and the aircraft has accelerated, the flight crew selects an appropriate thrust setting and navigation mode on the MCP.

Conditions: To be performed in a B-757/767 aircraft or flight simulator aligned on the appropriate heading and maintaining a target minimum airspeed of V_2 + 15 KIAS. No aircraft systems malfunctions exist; no abnormal weather conditions exist.

Standards: The PF should maintain assigned heading ±5° and a target minimum airspeed of V_2 + 15 KIAS. If conditions permit, the PF should reduce the pitch attitude at 1000 feet AGL and direct the PNF to retract the flaps on speed schedule as the aircraft accelerates. The flight crew should select an appropriate thrust setting and navigation mode on the MCP after the aircraft accelerates to V_{ref} 30 + 80. The flight crew should accomplish all required callouts and acknowledgements.

Supporting Proficiency Objective

Subtask: Initial Climb Procedures With Engine Failure

Performance: Prior to the engine failure, the PF maintains the assigned heading and a target minimum airspeed of V_2 + 15. The PNF monitors the aircraft systems and flight instruments and supports the PF as necessary. As the engine fails, the PF applies sufficient rudder and aileron pressure to maintain assigned heading, a positive rate of climb, and aircraft control. The PF maintains airspeed by applying forward pressure on the yoke to decrease pitch attitude to a target of 12.5°. The first pilot to recognize that a problem exists notifies the other pilot. The PNF examines the systems' instruments and notifies the PF which engine failed.

With the aircraft under positive control, the PF calls for and the PNF contacts ATC to declare an emergency and request a straight out departure if terrain permits. The PF also calls for and the PNF performs the memory items of the Engine Fire/Failure Checklist. When time permits, the flight crew discusses whether to return to attempt a landing at the airport or select an alternate airport. The PNF obtains latest weather report from ATC prior to making decision to return for an approach.

At 1000 feet AGL, the PF calls for and the PNF engages the Altitude Hold mode on the MCP. As the aircraft accelerates, the PF calls for and the PNF retracts the flaps on speed schedule. With the flaps retracted to their final setting, the PF calls for and the PNF engages FL Change on the MCP to climb to the assigned altitude.

Conditions: To be performed in a B-757/767 aircraft or flight simulator aligned on the appropriate heading and maintaining a target minimum airspeed of V_2 + 15 KIAS. One engine may fail at any point during the initial climb. No abnormal weather conditions exist.

Standards: The flight crew should not attempt to accomplish subordinate tasks until the PF has positive control of the aircraft. The PF should maintain aircraft heading ±5° and not allow the airspeed to deteriorate below V_2. The PNF should clearly communicate to the PF which engine has failed and support the PF as necessary to maintain aircraft control. Only when the aircraft is under control, should the PF direct the PNF to communicate with ATC and initiate the memory items of the Engine Fire/Failure Checklist. The PF should initiate aircraft reconfiguration as close to 1000 feet AGL as practicable. The appropriate final flap setting will depend on the flight crew's intentions. If the crew intends to return to the same airport to attempt another approach, then Flaps 5 may be the most efficient setting. However, if the flight crew intends to proceed to another airport, they should retract the flaps completely. The PF should remain cognizant that Auto Throttles are not available during single-engine operation. The flight crew should accomplish all required callouts and acknowledgements.

Terminal Proficiency Objective

Task: Instrument Landing System Approach

Performance: Perform an ILS approach to include three subtasks: Approach Preparation, Approach Transition, and Final Approach.

Conditions: The ILS Approach is to be performed in a B-757/767 aircraft or flight simulator under weather conditions that justify either a Category I, II, or III approach. Category I and II approaches may be accomplished with all aircraft systems normal or with one engine inoperative. Category III approaches can be accomplished only with all aircraft systems normal. The Captain must serve as the PF for all Category II and III approaches.

Standards: Performance standards for the subtasks Approach Preparation and Approach Transition are the same regardless of the conditions. However, performance standards for the subtask Final Approach are affected by the condition single-engine operation. Therefore, separate SPOs exist for the subtask Final Approach for normal conditions and for single-engine operation. Flight crews should be able to meet the performance standards specified in each of these supporting proficiency objectives.

Supporting Proficiency Objective

Subtask: Approach Preparation

Performance: If not accomplished previously, the flight crew contacts ATC or ATIS to verify the weather conditions that exist on the runway to determine whether to start the approach, what type of approach is required, and which pilot will serve as PF. The Captain directs the pilot that will not fly the approach to review the approach materials, tune and identify the radios, set the altimeter and airspeed bugs, and program the approach into the FMS if time permits. When completed, the pilot reviewing the approach notifies the other pilot and the crew conducts a positive exchange of controls. The pilot who will fly the approach reviews the approach materials, verifies the settings of the altimeter and airspeed bugs, and conducts the Approach Briefing. The Approach Briefing should include preparation for any abnormal conditions that may be anticipated. The PF calls for and the PNF completes the Approach Checklist.

Conditions: To be performed in a B-757/767 aircraft or flight simulator with weather conditions justifying either a Category I, II, or III approach and all aircraft systems normal. The Autopilot is engaged.

Standards: The flight crew should accurately interpret the weather conditions to identify the category approach required, which pilot will serve as PF, and the appropriate decision altitude and minimums. When transferring controls, the pilot relinquishing control should use the MCP to brief the receiving pilot about the current status of the aircraft and both pilots should clearly acknowledge who has control of the aircraft. The flight crew should attempt to complete approach preparation prior to receiving vectors to intercept the final approach course. The flight crew should accomplish all required callouts and acknowledgements.

Supporting Proficiency Objective

Subtask: Approach Transition

Performance: If the aircraft is not configured with Flaps 5, then the PF calls for and the PNF selects Flaps 5. The flight crew follows ATC vectors to intercept the final approach course. The PNF monitors the aircraft systems and flight instruments. The PNF notifies the PF of the first positive movement of the localizer and glide slope indicators. When ATC clears the aircraft for the approach, the PF selects Approach Mode on the MCP. The PF directs the PNF to extend the flaps on schedule and to extend the landing gear. The PNF extends the flaps, updates the airspeed on the MCP, extends the landing gear, and notifies the PF of each flap setting and when the landing gear are down. The PNF notifies the PF when the LOC annunciator changes from white to green and the previous roll mode disengages. The PNF notifies the PF when the GS annunciator changes from white to green and the previous pitch mode disengages. After the localizer and glideslope have been captured, the PF directs the PNF to set the missed approach altitude on the MCP. The PNF notifies ATC when the aircraft is at the final approach fix.

Conditions: To be performed in a B-757/767 aircraft or flight simulator with weather conditions justifying either a Category I, II, or III approach and all aircraft systems normal. The Autopilot is engaged and the Approach Checklist has been completed.

Standards: The flight crew should comply with all ATC instructions to intercept the final approach course. If the crew is not ready for the approach, then the PF should request additional time from ATC. The flight crew should attempt to configure the aircraft for landing prior to crossing the final approach fix. The flight crew should accomplish all required callouts and acknowledgements.

Supporting Proficiency Objective

Subtask: Final Approach

Performance: The flight crew completes the Before Landing Checklist. If Autoland is to be used, both pilots monitor the status and performance of the Autoland system. The PNF monitors the flight instruments and notifies the PF of any significant deviations from the desired flight path, airspeed, or descent rate. The PF monitors alternately the flight instruments and the outside environment to visually detect the runway environment. The PNF performs the required altitude or height callouts when indicated by the barometric and radio altimeters. Either pilot should clearly notify the other when he or she visually detects the runway environment. As the aircraft descends to Decision Altitude on Category I and II approaches, the PF decides whether to continue or terminate the approach. When the PF decides to continue the approach, the PF either continues with the Autoland or transitions to a visual approach and landing.

Conditions: To be performed in a B-757/767 aircraft or flight simulator with weather conditions justifying either a Category I, II, or III approach and all aircraft systems normal. The aircraft is configured for landing at the final approach fix with both the localizer and the glideslope captured.

Standards: Both pilots should monitor the status and performance of the autoflight system and remain ready to assume control of the aircraft if necessary. The PNF should accomplish the altitude or height callouts appropriate for the approach category. The first pilot to visually detect the runway environment should clearly notify the other pilot. The PF should clearly verbalize his or her decision to continue or terminate the approach.

Supporting Proficiency Objective

Subtask: Single-Engine Final Approach

Performance: The flight crew completes the Before Landing Checklist. If not accomplished previously, the flight crew should coordinate with ATC to plan a straight out departure in the event of a missed approach. The PF maintains the appropriate airspeed by manually controlling the throttle for the operative engine. If Autoland is to be used, both pilots monitor the status and performance of the Autoland system. The PNF monitors the flight instruments and notifies the PF of any significant deviations from the desired flight path, airspeed, or descent rate. The PF monitors alternately the flight instruments and the outside environment to visually detect the runway environment. The PNF performs the required altitude or height callouts when indicated by the barometric and radio altimeters. Either pilot should clearly notify the other when he or she visually detects the runway environment. As the aircraft descends to Decision Altitude on Category I and II approaches, the PF decides whether to continue or terminate the approach. When the PF decides to continue the approach, the PF either continues with the Autoland or transitions to a visual approach and landing.

Conditions: To be performed in a B-757/767 aircraft or flight simulator with weather conditions justifying either a Category I or II approach and one engine operating. The aircraft is configured for landing at the final approach fix with both the localizer and the glideslope captured.

Standards: The flight crew should remain cognizant that Autothrottles are not available during single-engine operations. The PF should maintain the appropriate airspeed ±5 KIAS. The PF should apply rudder to maintain aircraft trim. The PNF should accomplish the altitude or height callouts appropriate for the approach category. The first pilot to visually detect the runway environment should clearly notify the other pilot. The PF should clearly verbalize his or her decision to continue or terminate the approach.

Terminal Proficiency Objective

Task: Missed Approach

Performance: Perform Missed Approach Task to include two subtasks: Missed Approach Initiation and Missed Approach Initial Climb.

Conditions: Missed Approach may be accomplished under several condition sets that affect either the performance or the performance standards for the individual subtasks. These conditions exist throughout the entire task and include normal weather, single-engine operation, and flap or landing gear malfunctions.

Standards: Performance standards for the subtasks Missed Approach Initiation and Missed Approach Initial Climb are determined by the existing conditions. The existing conditions alter significantly the performance or the performance standards for the subtasks. Therefore, separate supporting proficiency objectives exist for each subtask that is affected by each set of conditions. Flight crews should be able to meet the performance standards specified in each of these supporting proficiency objectives.

Note: This Missed Approach TPO is written for coupled approaches. Hand flown approaches require significantly different performance by the flight crew and may require that a separate SPO be written.

Supporting Proficiency Objective

Subtask: Missed Approach Initiation

Performance: The PF evaluates the existing conditions and rapidly decides whether to terminate or to continue the approach. After deciding to terminate the approach, the PF quickly and clearly announces the decision to the PNF. Simultaneously, the PF activates the Go-Around mode by toggling the Go-Around switch on the throttles and directs the PNF to retract the flaps to the Missed Approach flap setting. The PF verifies on the ADI that the Missed Approach mode is active. The PNF sets the flaps, notifies the PF of the flap setting, monitors the flight instruments, and notifies the PF when the barometric altimeter indicates a positive rate of climb. When the aircraft is climbing, the PF calls for and the PNF retracts the landing gear. When the aircraft is stabilized and climbing, the PNF notifies ATC that the crew terminated the approach.

Conditions: To be performed in a B-757/767 aircraft or flight simulator with normal weather and no aircraft systems malfunctions. The aircraft is aligned with the final approach course of the runway and is coupled on the final segment of an ILS approach. The flight crew has completed the Before Landing Check.

Standards: The PF should quickly evaluate the conditions and immediately notify the PNF of the decision to terminate the approach. Simultaneous with the announcement, the PF should activate the Go Around mode and direct the PNF to retract the flaps to the missed approach setting. The PF should be alert that the automation engages correctly recognize any conditions indicating improper functioning of the Go Around mode. The crew should accomplish all call outs and acknowledgements required for retracting the flaps and the landing gear. After the aircraft is stabilized and climbing, the PNF should notify ATC that the crew terminated the approach.

Supporting Proficiency Objective

Subtask: Missed Approach Initial Climb

Performance: As the aircraft approaches 1000 feet AGL, the PF decreases the pitch attitude to approximately 10 degrees and directs the PNF to select Flight Level Change on the MCP, to set V_{ref} 30 + 80 on the Command Airspeed Bug, and to select climb power. The PF calls for and the PNF retracts the flaps on schedule. Once the flaps have been retracted and the aircraft has accelerated, the flight crew selects an appropriate thrust setting and navigation mode on the MCP.

Conditions: To be performed in a B-757/767 aircraft or flight simulator with go-around power applied, the landing gear retracted, and no systems malfunctions.

Standards: The crew should initiate this subtask as close to 1000 feet AGL as possible. The crew should comply with the speed schedule for retracting the flaps and accomplish all callouts and acknowledgements required for retracting the flaps. The PNF verifies the position of the flap handle and the Flap Position Indicator. The appropriate final flap setting will depend on the flight crew's intentions. If the crew intends to return to the same airport to attempt another approach, then Flaps 5 may be the most efficient setting. However, if the flight crew intends to proceed to another airport, they should retract the flaps completely.

Supporting Proficiency Objective

Subtask: Single-Engine Missed Approach Initiation

Performance: The PF evaluates the existing conditions and rapidly decides whether to initiate a Missed Approach or to continue the approach. After deciding to terminate the approach, the PF quickly and clearly announces the decision to the PNF. Simultaneously, the PF activates the Go-Around mode by toggling the Go Around switch on the throttles, manually applies Go Around power, and directs the PNF to retract the flaps to the Missed Approach flap setting. The PF verifies on the ADI that the Missed Approach mode is active. The PNF sets the flaps, notifies the PF of the flap setting, monitors the flight instruments, and notifies the PF when the barometric altimeter indicates a positive rate of climb. When the aircraft is climbing, the PF calls for and the PNF retracts the landing gear. When the aircraft is stabilized and climbing, the PNF notifies ATC that the crew terminated the approach.

Conditions: To be performed in a B-757/767 aircraft or flight simulator with normal weather and one engine operating. The aircraft is aligned with the final approach course of the runway and is coupled on the final segment of an ILS approach. The flight crew has completed the Before Landing Check.

Standards: The PF should quickly evaluate the conditions and immediately notify the PNF of the decision to terminate the approach. Simultaneous with the announcement, the PF should activate the Go Around mode, smoothly apply Go Around power to the operating engine, and direct the PNF to retract the flaps to the missed approach setting. The PF should remain cognizant that Auto Throttle mode is not available. The PF should be alert that the automation engages correctly and recognize any conditions indicating improper functioning of the Go Around mode. The crew should accomplish all call outs and acknowledgements required for retracting the flaps and the landing gear. After the aircraft is stabilized and climbing, the PNF should notify ATC that the crew terminated the approach.

Supporting Proficiency Objective

Subtask: Single-Engine Missed Approach Initial Climb

Performance: As the aircraft approaches 1000 feet AGL, the PF directs the PNF to select Altitude Hold on the MCP. As the PNF selects Altitude Hold, the PF must make the correct control inputs to maintain the aircraft stable and trim. As the aircraft accelerates, the PF directs the PNF to retract the flaps on schedule to their final setting. The PNF verifies the position of the flap handle and the Flap Position Indicator and notifies the PF after each change of flap setting. The final flap setting depends on whether the crew will attempt another approach at the same airport or move to another airport with better conditions. When the flaps are in their final setting and the aircraft has accelerated to the appropriate speed, the crew selects Flight Level Change on the MCP. Once the flaps have been retracted and the aircraft has accelerated, the flight crew selects an appropriate thrust setting and navigation mode on the MCP.

Conditions: To be performed in a B-757/767 aircraft or flight normal weather and one engine operating. All other aircraft systems are normal. The aircraft is climbing and has go-around power applied. The landing gear are retracted and the flaps are set at the missed approach setting.

Standards: The crew should initiate this subtask as close to 1000 feet AGL as possible. However, if either crew member must attend to a higher priority task, then this subtask may be delayed until both crew members can focus their attention on this subtask. The PF should apply appropriate rudder to prevent heading deviations as the crew changes MCP modes. The crew should comply with the speed schedule for retracting the flaps and accomplish all callouts and acknowledgements required for retracting the flaps. The PNF verifies the position of the flap handle and the Flap Position Indicator. The appropriate final flap setting will depend on the flight crew's intentions. If the crew intends to return to the same airport to attempt another approach, then Flaps 5 may be the most efficient setting. However, if the flight crew intends to proceed to another airport, they should retract the flaps completely.

APPENDIX D

INSTRUCTIONS FOR USE OF PERFORMANCE MEASUREMENT INSTRUMENT

Flight Crew Performance Assessment

Instructions

Assessing flight crew performance using this instrument requires that you wear two hats. The first will be the hat of a data recorder. In the first section of the instrument, you will record specific behaviors that you observe or do not observe. The form provides a list of these target behaviors for each maneuver. The second hat will be as an evaluator. For each maneuver, the form provides a rating scale for you to assess the overall performance of the flight crew on that maneuver. Details describing how you should perform each of these tasks are discussed below. To construct this performance assessment form we have made several assumptions. Some of these are described below.

1. We have focused on recurrent training for the B-757/767 aircraft. Therefore, the pilots and crews are proficient and experienced aviators. The majority of performance that you will observe will be satisfactory.

2. Although most performance will exceed standards set by the FAA, you will observe variability in performance between the different crews and pilots. That is, some crews will be better than others.

3. Performance on some tasks by some pilots will not be acceptable and will require repetition.

4. Specific pilot or crew behaviors that represent CRM skills can be identified for each task and these behaviors directly affect technical skills. CRM errors are often manifested in subsequent technical performance.

5. The instrument is designed for the one LOFT scenario, 7609. It does not allow for all possible contingencies but provides only for those conditions that are prescribed by the scenario script.

Behavioral Markers

We have identified specific behavioral markers of interest for each of the tasks included in the first leg of LOFT 7609. Your task is to indicate which markers you observe in each video tape. You will do this by checking the appropriate box only for those markers you do observe. We have attempted to list the markers in chronological order as they occur in the task. You should try to mark the form in real time. That is, complete the form as the crew performs the task. Unless absolutely necessary, do not wait until the end of the video tape to complete the form.

After the crew has completed each maneuver and you have marked the boxes, please provide a rating of overall performance for each pilot as well as the overall as a team. When providing these ratings you should answer two questions. First, was the performance satisfactory or not. If not, fill in the appropriate box under "UNSAT." If you thought performance was satisfactory, then provide a rating of how proficient for the appropriate crew member. Descriptions of the rating scale are provided at the end of this document. It is important to provide a rating for the crew as a whole. In addition, we realize that the two crew members may not be equally proficient. Therefore, we would like to get a measure of the proficiency of each crew member.

Bear in mind that these markers do not represent skills that a crew or pilot should demonstrate for Pass/Fail assessments. Our intention is to distinguish those skills exhibited by the best crews. In addition, not all of these markers should be observed all of the time. Their absence should be noted on the form only when relevant. For example, we are not interested in determining how well the autoflight system can hold a course. Thus, heading control is of interest only when the crew is hand flying the aircraft.

The following paragraphs provide details for observing the behavioral markers in each of the tasks on the assessment form.

Takeoff

The Takeoff Task begins as the crew taxis onto the active runway and ends as the crew changes the mode of the MCP after reconfiguring the aircraft at 1000 feet and accelerating. The Departure Briefing plays an important role throughout any flight. Therefore, we want to assess the crew's departure briefing but do not wish to include all of the prestart and taxi tasks. Thus, for the purposes of this study, we have included the Departure Briefing in the Takeoff Task. Descriptions of each of the behavioral markers for Takeoff follows.

1. Departure Brief has three components:

    a. Complete: The briefing should include all necessary items including those listed and those that should be anticipated.

    b. Eye Contact: The Captain should make eye contact with the FO when conducting the briefing. The Captain should make an honest attempt at conducting an effective briefing.

    c. PNF Contribute: The brief includes both members of the flight crew. Thus, the PNF should either add any items he or she thinks relevant or ask questions.

2. Complete To Brief: The PF should conduct the Takeoff Brief, include all required items, and the PNF should acknowledge the brief.

3. Complete All Callouts: Includes all callouts and acknowledgements required for the takeoff roll, rotation, and initial climb. These include: the call for N_1 EPR, 80 KIAS, Throttle Hold, Engines Check, V_1, V_r, V_2, Positive rate, and Gear Up.

4. Maintain Centerline: Throughout the takeoff roll, the PF should maintain aircraft alignment with the runway centerline ± 15 feet.

5. PNF Monitors Instruments: Throughout the takeoff roll, rotation, and initial climb, the PNF should monitor the engine and flight instruments. To some degree, whether the PNF is monitoring is a subjective call by the instructor. You cannot determine what the PNF is attending to. But, often you can tell what he or she is not attending to. This marker is to indicate when you are sure that the PNF is not monitoring the instruments. If the PNF's eyes appear to be on the instruments, then you must assume that he or she is monitoring them. However, when the PNF misses callouts or is clearly distracted by something else such as digging in his or her flight bag, you can assume that the PNF is not monitoring the instruments. The primary function of the PNF at this point is to support the PF if needed. The PNF cannot do this if he or she does not remain aware of the situation.

6. Rotation rate: The PF should rotate the aircraft at approximately 3° per second.

7. Pitch Attitude: The PF should use a target pitch attitude of 15°, not to exceed 18°. The PF should adjust the pitch attitude to achieve a target airspeed of $V_2 + 15$.

8. PF announces problem: The first crew member to detect the problem should notify the other that something is amiss. Typically, this would be the PF noticing that the aircraft requires unusual control inputs.

9. PNF confirms problem: The PNF should confirm the exact nature of the problem in very specific and succinct terms. For example, "The left engine has failed."

10. Heading control: As the engine fails, the PF should maintain heading control ±10° of the desired heading. If the heading deviates, the PF should take steps immediately to correct the deviation.

11. Roll control: As the engine fails, the PF should immediately apply rudder and aileron inputs to maintain aircraft stability.

12. Adjust pitch attitude: As the engine fails, the PF should reduce the pitch attitude to a target of 12.5° to maintain airspeed $\geq V_2$.

13. PNF supports PF: Through the engine failure, the PNF's principal responsibility is to support the PF as needed. This could be done verbally or in the worst cases by taking control. Again, this is a subjective call by the instructor. The question is whether the PNF provided support for the PF when it was needed. In some cases, the PF will not need any help or coaching.

14. Under control at 1000 feet: SAL procedures require that the crew reconfigure the aircraft at 1000 feet AGL. However, with the abnormal condition, the crew should accomplish this only when they are ready. The crew's first responsibility is to establish and maintain aircraft control. They should not attempt to reconfigure the aircraft until they have positive control. The instructor must judge whether the crew is ready to reconfigure when they begin.

15. Cleanup at 1000 feet: Did the crew begin to reconfigure the aircraft at or near 1000 feet AGL?

16. Use Altitude Hold: Did the crew level the aircraft to reconfigure by engaging the Altitude Hold Mode on the MCP?

17. Callouts for Flaps: The crew should accomplish all of the callouts, acknowledgements, and verifications to retract the flaps to the desired position.

18. Go to Flaps 5: The crew may immediately recognize that they will remain in the traffic pattern and return to Cincinnati. In this case many crews may choose to use Flaps 5 setting and remain at a lower airspeed.

Category 2 ILS Approach

The ILS approach begins as the crew calls for the Descent Checklist and ends when the Captain terminates the approach.

1. Descent Checklist: Did the crew call for and complete the descent checklist?

2. Engage Autopilot: The crew may have engaged the autopilot prior to beginning the approach. If not, they should engage it early on the downwind.

3. Check weather: If the crew have not checked the weather after the takeoff, then they should obtain the weather prior to committing to the approach.

4. Determine approach category: The crew should consider the weather and accurately determine the category of approach they will attempt. The crew should decide for the Captain to fly the approach.

5. MCP to exchange: On the downwind, the crew will need to exchange the controls so that each crew member can review and prepare for the approach. During these exchanges, the PF should use the MCP to brief the PNF about the status of the airplane. This is a means of achieving a shared situation awareness.

6. PNF calls complete: The PNF notifies the PF when he or she has completed all tasks preparing for the approach.

7. Verify settings: The PF verifies all of the bug and altimeter settings as he or she reviews and briefs the approach. There are several ways that this may be accomplished. The question is whether the PF verifies what the PNF set up.

8. Approach Briefing: The Captain conducts a thorough Approach Briefing.

9. Coordinate missed approach: With a single engine, the crew may decide that a straight out missed approach would be better. They might coordinate the straight out miss with ATC.

10. Workload distribution: The downwind provides an opportunity for the crew to distribute their workload. The issue is whether the crew prepared for the approach throughout the downwind or did not manage their workload on the downwind. If they had relatively light work periods followed by intense work periods where the crew appeared rushed or behind, the crew did not distribute their workload.

11. Communicate with passengers and company: Did the crew communicate with the company and the passengers?

12. Ample time: The marker refers to whether the crew became rushed or got behind on the downwind. If the crew were rushed, then this block should be marked.

13. Request more time: If the crew needed more time to prepare for the approach, did they request additional time from ATC?

14. PNF monitors: After ATC clears the flight for the approach, the PNF should monitor the flight and systems instruments. Again, this is a judgment of the instructor. The box should be marked if it is obvious that the PNF is not attending to the instruments.

15. PNF callouts: The PNF should notify the PF at the first movement of the glideslope and localizer indicators and when these modes are captured.

16. Configured at FAF: Is the aircraft fully configured for the approach as it crosses the final approach fix?

17. PF monitors throttles: The PF should monitor and adjust the throttles throughout the approach. The PF is monitoring the throttles if he or she has one hand on them.

18. PF monitors controls: This marker is an attempt to assess the PF's situation awareness. The aware PF tracks the progress of the autoflight system along the approach. The box should be marked if there is no evidence that the PF is following through on the controls (yoke).

Missed Approach

In this scenario, ATC will force the crew to execute a missed approach. The Captain will be the PF. The maneuver begins as the Captain announces the missed approach and ends as the crew changes the mode of the MCP after reconfiguring the aircraft at 1000 feet and accelerating. Each of the markers are described below.

1. PF announce go-around: The PF should clearly announce the decision to go-around.

2. Intend to use autopilot: Did the crew intend to use the automation to fly the missed approach or did they intend to hand fly it? The box should be marked if they did intend to use the automation.

3. Engage TOGA: Did the PF engage the go-around mode when initiating the missed approach?

4. Apply power: Did the PF apply go-around power when initiating the missed approach?

5. Hand fly go-around: Did the PF hand fly the missed approach?

6. Heading control: If the PF hand flew the missed approach, did he or she maintain the desired heading ± 10°?

7. Roll control: If the PF hand flew the missed approach, did he or she maintain adequate roll control?

8. Rudder Trim: If the PF hand flew the missed approach, did he or she maintain rudder trim adequately?

9. Flaps 5: Did the crew reconfigure the aircraft to Flaps 5 as they initiated the missed approach?

10. Complete callouts: Did the crew complete all of the required callouts and acknowledgements as they initiated the missed approach? These include retracting the flaps to Flaps 5, positive rate, and gear up.

11. Clean up @ 1000 feet: Did the crew begin to reconfigure the aircraft at or near 1000 feet AGL?

12. Altitude Hold: To reconfigure the aircraft during single-engine operations, the crew should level the aircraft by engaging Altitude Hold. Some crews may use the normal operations technique by engaging Flight Level Change.

13. Aircraft control @ cleanup: SAL procedures require that the crew reconfigure the aircraft at 1000 feet AGL. However, with the abnormal condition, the crew should accomplish this only when they are ready. The crew's first responsibility is to establish and maintain aircraft control. They should not attempt to reconfigure the aircraft until they have positive control. The instructor must judge whether the crew are ready to reconfigure when they begin.

14. Anticipate trim requirement: When the crew engages Altitude Hold, the autopilot will stop applying rudder input. The PF should anticipate the need to apply rudder and be ready as they engage Altitude Hold.

15. PNF monitors: As the PF initiates the missed approach, the PNF should monitor the systems and flight instruments. The box should be marked if the PNF obviously did not monitor the instruments.

16. PNF supports: The PNF should support the PF as needed. Similar to when the engine first failed, this may be verbally or even helping on the controls. The box should be checked if the PNF failed to provide support as needed during the missed approach.

Enroute

This task serves as a "catch all" and includes two unrelated components that do not appear to fit elsewhere: Turn to Downwind and Rerouting to Alternate Airport. The Turn to Downwind does not technically belong as part of the Takeoff or the ILS Approach. It is a transition period between the two tasks that is created by the events of this particular LOFT scenario. However, we have found that several interesting indicators of crew performance may occur during this time. So, we have included it here in hopes that we will find a better home for it later.

We included Rerouting to the Alternate Airport in an attempt to include a task that is relatively normal and does not involve any systems malfunctions. Within the constraints of this LOFT, this comes closest to normal operations.

Turn to downwind:

1. Decision to return:

   a. Timely: Is the decision to return made rapidly? Some pilots indicate they know they will return immediately after the engine problem develops.

   b. Discuss with FO: Does the Captain discuss the decision with the FO?

   c. Consider weather: Does the crew consider the weather before committing to an approach? This may be accomplished early in the downwind segment.

   d. Announced: Does the Captain announce the decision to return to Cincinnati?

2. Hand fly: Does the PF choose to hand fly the aircraft through the turn? The box should be marked if the PF engages the autopilot before getting established on downwind.

3. Trim rudder: Does the PF manually provide trim as needed?

4. Complete checklists: Does the crew call for and complete the After Takeoff and Engine Fire/Failure Checklists during this segment of the flight?

5. Communicate: Does the crew communicate with the passengers and the company during this segment of the flight?

6. Engage Autopilot on downwind: Does the crew engage the autopilot after getting established on downwind?

Rerouting to Alternate Airport:

1. Decision to divert:

   a. Timely: Is the decision to divert made rapidly? Some pilots indicate they know they will divert immediately after they initiate the missed approach.

   b. Discuss with FO: Does the Captain discuss the decision with the FO?

   c. Consider factors: Does the crew consider all relevant factors before making the decision? These include the weather, maintenance, passengers, the company.

   d. Announced: Does the Captain announce the decision to the FO to ensure both crew members understand the course of action?

2. Use single-engine cruise information: Does the crew use the single-engine cruise information provided by the FMS?

3. Use FMS navigation: Does the crew use LNAV and VNAV functions to navigate to the alternate?

4. Announce programming complete: When finished programming the FMS, the PNF notify the PF that LNAV and VNAV are available?

Crew Performance Rating Scale

In addition to recording specific behaviors that you observe, you will assess the crew's proficiency on each maneuvers. Bear in mind that you will be rating the proficiency of the crew and not an individual pilot. As the crew completes each maneuver, you will assess the crew's overall performance on that maneuver. Your assessment will be subjective and based on your previous experiences as pilot, instructor, and evaluator. That is, you will judge the crew's proficiency based on what you have seen other crews do in the past.

The Crew Performance Rating Scale is divided into two sections and will require that you make two decisions. The first is whether crew performance was satisfactory. The second is to judge the level of crew proficiency if their performance was satisfactory. Again, both of these are subjective judgments on your part. We are relying on your expertise and experience to provide judgments of pilot and crew performance.

Immediately after the maneuver, you will judge whether the crew performed the maneuver satisfactorily or not. If for any reason you felt their performance was unsatisfactory, then mark the appropriate box, note on the back your reasons for this rating, and move on to the next maneuver.

If you judge crew performance to be satisfactory, then you will rate the crew's level of proficiency using a five-point scale. Since all of the pilots you will see are current SAL pilots in recurrent training, we anticipate that the large majority of performance will be satisfactory. But, pilots in recurrent training vary considerably in their proficiency. The purpose of this scale is to capture some of that variability. In addition, we recognize that there are often many acceptable ways to accomplish the same task. Many of the differences among pilots' performances are a matter of personal technique. However, we also recognize that some techniques are better than others and that pilots vary in their abilities and performance.

The anchors for the five-point rating scale are described below. Note that only three anchors are presented on the form. This was due to a lack of space. However, you should freely use any of the five boxes that you think are appropriate.

<u>Acceptable</u>: This crew accomplishes the maneuver safely, but their performance is barely acceptable. These are the least proficient of SAL's pilots. You might recognize that these pilots need additional training and you might not be totally comfortable with them flying your spouse and kids. The crew has difficulty remembering procedures; they are late responding to the situation; they are always behind; they have difficulty managing the aircraft and fail to use the automation effectively; the pilots operate as individuals and not as a team.

<u>Below Average</u>: This crew is less proficient than the average SAL crew and needs considerable improvement. They probably have some difficulty staying ahead of the aircraft in normal and abnormal conditions. The crew does not perform well as a team; one pilot

may be strong and carry the entire crew. The crew may have sufficient technical skills but fail to accurately assess the situation or anticipate future needs and events.

Average: This is the average SAL crew. The crew is good and safe, but you can see some room for improvement in both their technical skills and their CRM skills. The crew is friendly and work well together; they stay ahead of the aircraft during normal operations, but they are slow on the procedures that they perform infrequently.

Above Average: This crew is better than the average SAL crew, but not yet experts. They have good technical skills and stay ahead of the aircraft most of the time. They might communicate well and share the same situation assessments. They may not manage their time or anticipate future events.

Expert: This crew is clearly expert; they rank among the best that you have seen. They conduct the maneuver without any problems. They handle the aircraft smoothly and utilize the automation efficiently. The crew members communicate effectively with one another; they anticipate each other's needs as well as future events; they manage their time well; and they share an accurate estimate of the situation.

☆ U. S. GOVERNMENT PRINTING OFFICE: 1993--701-889 (6170)