Crew Familiarity: Operational Experience, Non-Technical Performance, and Error Management

Matthew J. W. Thomas and Renée M. Petrilli

Thomas MJW, Petrilli RM. ***Crew familiarity: operational experience, non-technical performance, and error management*****. Aviat Space Environ Med 2006; 77:41–5.**

Introduction: Crew familiarity, in terms of having recent operational experience together as a crew, is seen as an important safety-related variable. However, little evidence exists to unpack the underlying processes with respect to familiarity. This study investigated the relationships between crew familiarity, non-technical performance, and error management. ***Method:*** Data were collected during normal line operations at a commercial airline by observers using a methodology based on the Line Operations Safety Audit (LOSA). A total of 154 flights were analyzed, 31% of which were operated by an unfamiliar crew, with 69% operated by a familiar crew. ***Results:*** The rate of error occurrence was found to be higher for unfamiliar crews, and these crews were found to make different types of errors when compared with familiar crews. However, with respect to the management of error events, no significant difference was found between unfamiliar and familiar crews. No significant effect of crew familiarity was found with respect to crews' non-technical performance. However, strong correlations were found between crews' non-technical performance and the management of errors. ***Discussion:*** The findings indicate that crew familiarity, in terms of whether a crew has flown together recently or not, has little operational significance with respect to the management of error events during normal line operations. Accordingly, the suggestion that unfamiliar crews operate at a higher level of safety-related risk was not supported. Non-technical performance appears to be a stronger driver of effective error management than crew familiarity, and is highlighted as a focus for further investigation and intervention.

Keywords: crew resource management, human error, rostering.

THE FAMILIARITY of flight crew, in terms of having recent operational experience together as a crew, has been the focus of speculation with respect to either enhancing or impairing operational performance and safety. Limited empirical investigation of crew familiarity has taken place, and significant conflicting evidence exists with respect to the relationship between crew familiarity and safety. In a major simulator-based study of commercial airline short-haul crew, it was found that flight crew with recent operating experience together performed better on a number of dimensions relevant to flight safety than crews who had not flown together recently. In this study, crew familiarity was seen to improve crew communication, and specifically the willingness of crewmembers to exchange information. Similarly, crews with recent operating experience together made less operationally significant errors than crews who had not (3).

Typically, the beneficial elements of crew familiarity have been explained in terms of crewmembers developing an appreciation of each other's operating style. After operating several flights together, crewmembers increase their knowledge and awareness of individual preferences for interaction, and are able to tailor their own style to suit the needs of other crewmembers. Accordingly, "familiar" crews maximize the opportunity to convey operationally significant information through this process of tailoring communication to match the preferences of other crew. Analyses of accident data in commercial aviation have reinforced these research findings, with a recent controlled flight into terrain accident report highlighting that a lack of crew familiarity can contribute to crew failing to adequately monitor and communicate exceedences from critical flight parameters (11). Similarly, a meta-analysis of U.S. commercial aviation accidents between 1978 and 1990 in which crew actions were involved as a contributory factor found that in 73% of the 15 accidents for which data was available, the crew had not flown together before the day of the accident (10).

However, crew familiarity has also been implicated in the degradation of crew performance. For instance, even the two-person crew of the modern flight deck can be seen as susceptible to "groupthink," a term coined to describe the negative effects of team cohesion on effective decision making (7). According to the theory of groupthink, one of the implications of over-familiarity is the reduction in monitoring and criticism of others' actions. A number of recent commercial aviation accident and incident investigations have highlighted such negative processes in crews who had flown together frequently in a period leading up to an accident or incident. In one example, crew familiarity was determined to be a likely cause for an inadequate approach briefing, which then led to an attempted visual approach in poor weather and near controlled flight into terrain (1). Similarly, empirical research has found a lower error rate in unfamiliar military crews when com-

From the Centre for Applied Behavioural Science, University of South Australia, Woodville, Australia.

This manuscript was received for review in July 2005. It was accepted for publication in October 2005.

Address reprint requests to: Matthew J. W. Thomas, Ph.D., who is a Senior Research Fellow, Centre for Applied Behavioural Science, University of South Australia, GPO Box 2471, SA 5001, Australia; matthew.thomas@unisa.edu.au.

pared with "fixed" crews who operated together indefinitely, reinforcing the possibility that crew performance declines as a factor of over-familiarity (2). Complacency is, therefore, implicated as an important factor mediating the performance of over-familiar crews.

Within the current rostering practices of airline operations worldwide, crew will not normally fly together for an extended period of time. Typical rostering patterns in short-haul operations involve flying with the same crewmember for a 3- to 4-d trip, and flying several flight sectors per day. Within this rostering environment, flight crew are frequently operating with other crewmembers who they have not operated with before. Accordingly, crew familiarity, in terms of unfamiliar crew, has the potential to pose a significant operational risk if not effectively understood and managed. Over the last few decades, Crew Resource Management (CRM) training programs have focused on the development of specific skills in the "non-technical" areas of communication, situation awareness, problem solving, and stress management (4). These skills are now seen as critical operational countermeasures with respect to error occurrence and management, and accordingly have the potential to mitigate the effects of unfamiliarity on crew performance.

The primary objective of this study was to examine whether a relationship existed between crew familiarity and patterns of error occurrence and error management during normal flight operations. A secondary objective was to examine whether relationships existed between crew familiarity, the non-technical performance of crew, and crew error management behaviors. The study used an existing dataset of naturalistic observations of normal line operations.

METHODS

Participants

Data were collected during the observation of normal line operations within the narrow-body twin jet fleet of a single commercial airline. Of the 215 single flight sectors in the dataset, 61 were excluded from analysis due to irregularities in crew composition, such as crew changes mid-pattern due to illness, or the crew had flown together before, but not in the days directly prior to the observation flight. All flight sectors observed were flown by two crewmembers, with a captain operating in the left-hand seat and a first officer operating in the right-hand seat as per normal standard operating procedures.

All crew participated in the data collection process on a voluntary and anonymous basis. The data collection was funded by the airline involved in the study as part of an anonymous system-wide audit of flight operations. All crew were able to decline an observation, and no individual crewmembers were identified in the data collected.

Observational Methodology

Observational data were collected by expert trained observers using a methodology based on the Line Operations Safety Audit (LOSA) developed by the University of Texas and promulgated by the International Civil Aviation Organization (5,8). The observer sat in the jump seat of the airplane, between and behind the two crewmembers. The LOSA methodology provides a structured observational framework for the collection of data relating to crew performance. The methodology collects data relating to a crew's threat and error management behaviors. Within this framework, errors are defined as crew action or inaction that leads to a deviation from crew or organizational intentions or expectations, and are taken to be an unavoidable and ubiquitous aspect of normal operations (9).

Observers attended a four-day training session in the use of the observational methodology. This training included 2 d in the classroom learning the fundamentals of the observation process, a day of trial observations during normal flight operations, and a final day of calibration. Observer calibration and inter-rater reliability was established on the final day of training. Training was deemed complete when scores on the within-group inter-rater reliability statistic were consistently maintained above 0.70 for observers' ratings of video-based examples of normal flight operations (6).

Measures

Crew familiarity was recorded for the purposes of this study by the observer as a dichotomous categorical variable identifying whether the crew had flown together before the day of the observation or had not flown together before. This choice was driven in the first instance by the necessity to collect demographic data from observed crew as efficiently as possible during relatively short flight sectors. Secondly, there was a desire to record crew familiarity in terms that were operationally relevant within the context of current rostering practices.

Data pertaining to the occurrence and management of crew error were initially coded by observers in relation to four key variables: 1) error type; 2) error detection; 3) error response; and 4) error outcome. This coding framework yielded data in relation to the management of each individual error event. However, for the purposes of this study, the error management data for each flight sector were aggregated to form composite error management ratings that included: 1) the number of errors per sector; 2) the proportion of inconsequential errors per sector; 3) the proportion of effectively managed errors per sector; and 4) the number of undesired aircraft states per sector resulting from mismanaged errors.

Observers also rated crew performance with respect to a series of non-technical behavioral markers used to measure the effectiveness of crew performance in relation to key crew resource management skill dimensions. A total of 16 behavioral markers under the four categories of: 1) communication; 2) situation awareness; 3) task management; and 4) decision making were used. The individual markers and the coding strategies used in this type of data collection have been described in detail previously (12,13). The behavioral marker system has been designed to avoid potential problems of com-

TABLE I. DISTRIBUTION OF ERROR TYPES ACCORDING TO CREW FAMILIARITY

Error Type	Frequency (% of errors)	
	Unfamiliar	Familiar
Aircraft Handling	12.0	12.0
Flight Management Computer	7.3	5.4
Flight Controls	1.0	5.0
Mode Control Panel	13.0	15.1
Systems/Instruments/Radios	11.5	10.1
External Communication	11.5	5.7
Checklist	4.7	7.6
Standard Calls	16.1	21.1
Briefings	4.2	3.5
Monitoring and Cross-Checking	3.6	5.4
Paperwork	3.6	0.6
Airmanship	6.8	6.6
Takeoff and Landing Data	3.1	0.9
Other	1.6	0.9

N = 509.

mon method variance, where crews' error management outcomes can influence overall non-technical skill ratings. The system does this through the provision of clearly defined and differentiated markers that target specific categories of behaviors. To this end, ratings of non-technical skills remain as independent markers of specifically observed behaviors, rather than generic indicators of error management outcome or overall crew performance.

RESULTS

Of the 154 flight sectors in the dataset, 47 (31%) were operated by an unfamiliar crew who had not flown together before the day on which the flight observation took place, with the remaining 107 (69%) operated by a familiar crew who had operated together previously in the days prior to the observed flight. During the 154 flight sectors in the dataset, a total of 509 individual error events were observed, with an average rate of 3.31 errors (SD = 2.62) per flight sector. The average number of errors per flight sector operated by unfamiliar crews was 4.09 (SD = 2.90), compared with an average of 2.96 errors (SD = 2.42) per flight sector operated by familiar crews. From this data, error occurrence was found to be significantly higher for unfamiliar crews [F (1,152) = 6.220, p = 0.014]. The types of errors made by unfamiliar crews differed significantly when compared with those made by familiar crews [χ^2 (13,509) = 25.553, p = 0.020]. Unfamiliar crews were found to make proportionately less errors relating to flight controls, checklist use, standard calls, and monitoring and cross checking, and proportionately more errors relating to paperwork and external communication. Details of the distribution of error type for unfamiliar and familiar crews are presented in **Table I**.

With respect to the management of error events, no significant difference was found between unfamiliar and familiar crews. The overall average proportion of errors that were detected and actively managed to an inconsequential outcome by crews was 62.56% (SD = 33.32), with an average of 61.54% (SD = 32.84) for unfamiliar crews, compared with an average of 63.06% (SD = 33.72) for familiar crews [F (1,133) = 0.062, p = 0.803]. Similarly, the overall average proportion of inconsequential errors, including those that were not detected by the crews, was 91.81% (SD = 17.82), with an average of 91.43% (SD = 19.65) for unfamiliar crews, compared with an average of 92.00% (SD = 16.94) for familiar crews [F (1,133) = 0.030, p = 0.863]. While no significant differences were found in relation to the error management behaviors of familiar and unfamiliar crews, the overall higher rate of error occurrence for unfamiliar crews led to a slightly elevated incidence of undesired aircraft states, with unfamiliar crews having one undesired aircraft state per 5.88 flights compared with one undesired aircraft state per 7.14 flights for familiar crews. However, due to the small number of undesired states overall, this difference was not found to be statistically significant.

No significant effect of crew familiarity was found with respect to crews' non-technical performance scores on each of the 16 behavioral markers. However, consistently strong correlations were found between crews' non-technical performance and the crews' management of errors. This relationship was found to be most significant with respect to the interactive crew functions of communication, such as leadership/followership, assertiveness, and inquiry. Details of each of the behavioral markers and the correlation analyses are presented in **Table II**.

DISCUSSION

The results of this study indicate that crew familiarity, in terms of whether a crew has flown together recently or not, has little operational significance with respect to the management of error events during normal line operations. Specifically, crew familiarity was found to have no significant effect with respect to the detection and mitigation of error events. There was no evidence to support the suggestion that unfamiliar crews operated at a higher level of safety-related risk.

The overall rate of error occurrence was found to be

TABLE II. RELATIONSHIP BETWEEN PROPORTION OF EFFECTIVELY MANAGED ERRORS AND NON-TECHNICAL MARKERS

Non-Technical Marker	Correlation Coefficient
Communication Environment	0.278**
Leadership/Followership	0.356**
Inquiry	0.285**
Assertiveness	0.293**
Cooperation	0.145
Statement of Plans and Changes	0.247**
Vigilance	0.299**
Monitoring and Cross-Check	0.243**
Briefing and Planning	0.243**
Workload Management	0.077
Workload Assignment	0.276**
Automation Management	0.045
Fatigue and Stress Management	0.161
Contingency Planning	0.180*
Problem Identification	0.213*
Evaluation of Plans	0.183

* = $p < 0.05$; ** = $p < 0.01$.

higher for unfamiliar crews, with these crews creating proportionately more errors than familiar crews. While the difference between the "baseline" error rates of unfamiliar and familiar crews is an important finding, the finding of different patterns of error occurrence, particularly with respect to the types of error committed, is perhaps of greater operational relevance. The study found significantly different distributions of error phenotype (the operational manifestation of errors), with unfamiliar crews making proportionately more errors in the functions of aircraft paperwork and external communication. In contrast, unfamiliar crews made proportionately less errors relating to critical interactive crew functions such as monitoring and crosschecking, standard calls, and use of checklists. This finding highlights that while unfamiliar crews made significantly more errors overall, they made proportionately less errors in a number of core safety-critical functions.

The finding of significantly different distributions of error phenotype indicates the presence of behavioral differences between familiar and unfamiliar crews. While it is not possible to infer from these results the presence of systematic countermeasure use, the findings do generate the hypothesis that unfamiliar crews employ strategies to prevent the occurrence of error relating to critical interactive functions of the multi-crew operating environment, which in turn offer significant protection against safety-related risk. This is one possible explanation as to the mechanisms behind the slightly elevated error rate in unfamiliar crews, which might be an emergent feature of crews placing greater operational emphasis on critical multi-crew functions at the expense of less safety-critical processes.

The present study found no relationship evident between crew familiarity and the crews' scores on a set of 16 behavioral markers for non-technical (CRM) performance. This finding suggests that a crews' non-technical performance is resilient to any effects of unfamiliarity. The study did, however, find a significant relationship between the non-technical performance of crews and their management of error events. Accordingly, the results of this study indicate that the non-technical performance of a crew appears to be a stronger driver of effective error management than crew familiarity. As such, this study highlights an area for further empirical investigation with respect to safety-related intervention.

The finding of strong relationships between error management and non-technical performance provides further support for the effectiveness of non-technical skills as operational countermeasures, both in terms of error management, but also in relation to ameliorating any potential underlying effects of crew familiarity. While the correlation analyses presented in this study cannot support any real inference of causation, it is possible to hypothesize that CRM training programs and the standardization that detailed standard operating procedures bring to normal line operations reduce the potential influence of unfamiliarity on crew performance. Certainly, both CRM training and standard operating procedures have become cornerstones in efforts to enhance operational safety and performance within the commercial airline environment. In particular, the lower frequency of errors relating to the critical interactive crew functions such as monitoring and cross-checking, standard calls, and use of checklists, as well as the clear relationship between non-technical performance and effective error management found in this study, provide solid evidence to support the use of CRM and standard operating procedure countermeasures by unfamiliar crew. Much of the previous research that has indicated a significant effect of crew familiarity on performance was undertaken before CRM training programs had become fully embedded within mainstream airline culture. In contrast, the data analyzed in this study were collected in an environment that has undoubtedly benefited from several decades of CRM training programs, and where the development of crews' non-technical skills has become a regulated aspect of both initial and recurrent airline training programs. Furthermore, the data were collected from an airline that is characterized by a culture of strong adherence to standard operating procedures, low levels of intentional non-compliance errors, and a well-established CRM training program.

The present study was designed as an exploratory investigation of a naturalistic dataset, and accordingly the generalization of these results, particularly with respect to airlines with different organizational and national cultures, should be approached with caution. Similarly, the interpretation of the results is based on the assumption that error events during normal line operations are of the same class as the error events that occur as precursors to incidents or accidents. This assumption is in line with our current understanding of system safety, which is built on the belief that improving the management of "everyday" errors can contribute directly to a reduction in safety-related events such as incidents and accidents. However, it is acknowledged that further research is required to better differentiate these "everyday" errors from those that occur during "non-normal" operational conditions or aircraft states.

Given these caveats, the results of this study do raise significant questions with respect to existing thought in relation to the negative implications of crew familiarity. In summary, this study has reinforced the importance of non-technical performance as a critical component of safety in airline operations. It also suggests that effective non-technical performance may mitigate the potential effects an unfamiliar crew brings to commercial airline operations.

ACKNOWLEDGMENTS

Many thanks to the airline involved in the initial collection of data used in this study. The willingness of the airline and its crews to participate in the open collection of data during normal line operations demonstrates a level of reflexive self-evaluation essential to enhancing system safety. Similarly, many thanks to Captain Stan Prout for ongoing informative discussions, unrelated to this particular dataset, but nonetheless exploring the role of crew standardization as a systematic countermeasure against a range of factors such as crew unfamiliarity and fatigue.

REFERENCES

1. ATSB. AVRO 146-RJ70A VH-NJW, Norfolk Island Aerodrome, investigation report 199003515. Canberra, Australia: Australian Transport Safety Bureau; 1999.
2. Barker JM, Clothier CC, Woody JR, et al. Crew resource management: a simulator study comparing fixed versus formed crews. Aviat Space Environ Med 1996; 67:3–7.
3. Foushee HC, Lauber JK, Baetge MM, et al. Crew factors in flight operations III: the operational significance of exposure to short-haul air transport operations. Houston, TX: National Aeronautics and Space Administration; 1986. NASA Technical Memorandum 88322.
4. Helmreich RL, Wilhelm JA. Outcomes of crew resource management training. Int J Aviat Psychol 1991; 1:287–300.
5. ICAO. Line operations safety audit (LOSA). Montreal, Canada: International Civil Aviation Organisation; 2002.
6. James L, Demaree RG, Wolf G. Estimating within-group interrater reliability with and without response bias. J Appl Psychol 1984; 69:85–98.
7. Janis I. Victims of groupthink. Boston: Houghton Mifflin; 1972.
8. Klinect JR. LOSA searches for operational weaknesses while highlighting systemic strengths. International Civil Aviation Organisation (ICAO) Journal 2002; 57:8–9, 25.
9. Klinect JR, Wilhelm JA, Helmreich RL. Threat and error management: data from line operations safety audits. In: Jensen RS, ed. Proceedings of the tenth international symposium on aviation psychology. Columbus, OH: Ohio State University; 1999:683–8.
10. NTSB. A review of flightcrew-involved major accidents of U.S. air carriers, 1978 through 1990. Safety Study NTSB/SS-94/01. Washington, DC: National Transportation Safety Board; 1994: 40–2.
11. NTSB. Boeing 727–232 N497FE, Tallahassee, Florida, aircraft accident report NTSB/AAR-04/02 (PB2004–910402). Washington, DC: National Transport Safety Board; 2004:49–51.
12. Thomas MJW. Improving organisational safety through the integrated evaluation of operational and training performance: an adaptation of the line operations safety audit (LOSA) methodology. Human Factors and Aerospace Safety 2003; 3:25–45.
13. Thomas MJW. Predictors of threat and error management: identification of core non-technical skills and implications for training systems design. Int J Aviat Psychol 2004; 14:207–31.