A review of the benefits of aviation human factors training

Graham D. Edkins

Qantas Airways Limited
Corporate Safety Department
203 Coward Street Mascot NSW 2020 Australia
+ 61 2 9691 8095 (phone)
+ 61 2 9691 8833 (fax)
gedkins@qantas.com.au
A review of the benefits of aviation human factors training

Abstract

This paper reviews the available evidence for the benefits of aviation human factors training. Despite the proliferation of human factors training programs across the aviation industry since the 1980’s there are few published studies that demonstrate positive shifts in attitude or behavior following the introduction of such training. Those studies reporting benefits suffer from a number of methodological weaknesses including the failure to use control groups, lack of longitudinal evaluation and small sample sizes. Of significant concern is a lack of cost effectiveness data demonstrating a return on investment from human factors training. Recommendations for future research are made in the light of consolidating existing evidence on the commercial benefits of human factors training.

Introduction

Over the past twenty years, there has been an increasing recognition that human factors training can have a significant impact on safety, particularly within high-risk domains such as transport, mining and energy. The aviation industry has been at the forefront of this trend. Human error has been consistently identified as contributing to the majority of air crashes since the late 1970s (FAA, 1995; Johnson, 2000a; Salas, Prince, Bowers, Stout, Oser & Cannon-Bowers, 1999b). In most cases, these human errors occurred despite the operators’ technical competence. Highly trained people committed what seemed to be trivial errors. Usually these errors involved factors such as breakdown in communications, poor decision-making and failures in leadership (Helmreich, 1996) and in some cases resulted in tragic loss
of human life, as well as huge financial and public relations costs for those organisations involved.

To minimise human error, the aviation industry introduced a series of cockpit resource management training programs in the early 1980s. These resource management programs have typically raised awareness about human factors concepts, such as communication, decision making and teamwork, within the flight crew working environment. Airlines throughout the world have progressed from cockpit orientated training programs to Crew Resource Management (CRM) and Advanced Crew Resource Management (ACRM). Subsequently, these programs have expanded into a plethora of different forms across a wide variety of industries. For example, the air traffic control domain has adapted airline CRM training into a set of Team Resource Management programs (Andersen & Bove, 2000). Outside aviation, CRM-style training has been applied within the nuclear industry (Harrington & Kelso, 1991), the medical fraternity (Davies & Helmreich, 1996; Fisher, Phillips, & Mather, 2000; Howard, Gaba, Fish, Yang & Sarnquist, 1992), for maritime crews (Andersen, Soerensen, Weber & Soerensen, 1996; Bydorf, 1998), and in the offshore oil industry (Flin, 1995; Flin & O’Connor, 2001). Variants of CRM have also been used to enhance decision making and team skills training for emergency services personnel. Similarly, team based safety behaviour training, which is analogous to human factors training, has been incorporated into industries such as construction (eg. Raya, Bishop, & Qi, 1997). Regardless of industry application, these human factors training programs have similar foci on safety and human behaviour.

Despite the widespread use of human factors training programs, there is a lack of longitudinal studies examining their effectiveness. For example, do human factors training programs have
a significant impact on error management? Is there a relationship between human factors training and improved safety performance? The answers to these questions are particularly important for organisations questioning whether such programs are a sound investment in times when training budgets may be scarce.

The purpose of this paper is to critically review the current state of human factors training within the aviation industry with specific regard to: studies that claim a successful link between human factors training and improved safety performance; and programs that appear to demonstrate a return on investment. In particular, attention is paid to those studies reporting benefits from human factors training application in non-flight crew areas. It is hoped that this review may provide a precis of the status of human factors work on this subject as well as offer a useful resource for organisations to build a stronger human factors business case.

**An empirical analysis of human factors training**

While human factors training is becoming established across many industries, it remains most widespread in aviation. Successive generations of training programs have refined the concepts and added to their scope. However, from the outset, doubts have been expressed about the validity of programs. Initial doubters described the courses as “charm school” and irrelevant to aviation (Helmreich, 1996). Since then, program content has matured and the training methods have become increasingly sophisticated, but the doubters persist. Steps to overcome these doubters’ resistance have usually been fruitless. But do the doubters have the evidence on their side?
At first glance, the best method to evaluate human factors training programs would appear to be simple: rates of accidents. This approach seems logical given that the first programs were introduced to reduce the incidence of human error in accident rates. Yet despite twenty years of human factors training, human error continues to be involved in the majority of aviation accidents. Johnson (2000a) recently found that while CRM style courses have been used in many airlines, there has not been any noteworthy reduction in rates of incidents and accidents resulting from crew teamwork and communication failures.

However, using accident rates to evaluate human factors training is futile, due to significant changes over this period in the way aviation accidents are analysed. Aircraft safety occurrences are now investigated in much greater detail, with “multiple causality” now typically identifying a chain of events that leads to an accident (ICAO, 1992; Reason, 1992). Detailed systemic investigations almost inevitably lead to the identification of human error at one or more points along the chain. Such detailed analysis allows greater understanding of the causes of accidents, but it also means that the rate of accidents classified as “human error” has increased.

With the increasing sophistication of investigation techniques, accident rates cannot be used as a broad method of evaluating training success. In addition, because accident rates in commercial aviation are already low, the infrequency of events makes any detection in improvement difficult even with identical accident investigation techniques (Helmreich, Chidester, Foushee, Gregorich & Wilhelm, 1990; Salas et al., 1999a). Accident rates could only provide valid comparisons if pre-training and post-training accident rates are available, and if the methods of analysing cause and the error classification system are identical. Without such data, human factors training has to be evaluated in other ways.
However, there are examples of when human factors training has been cited as preventing accidents. For example, potential disasters with two United Airlines flights were averted and attributed to successful training in communication and teamwork (Salas, Rhodenizer, & Bowers, 2000). While such incidents are suggestive of benefits, anecdotal evidence is difficult to evaluate scientifically. Without rigorous controls, it cannot be determined if human factors training prevented the occurrence of accidents in these or other cases.

A review of the literature reveals that there are two approaches to evaluating the effectiveness of human factors training programs; attitudinal and behavioural.

**Attitudinal based methods of evaluating human factors training effectiveness**

In the early development of human factors training, emphasis was often placed on attitudes. This was typically true in both training and in evaluation. In the aviation industry, formative courses sought to modify trainee’s attitudes toward teamwork and other human performance issues. The effectiveness of these programs was usually evaluated by measuring employee attitudes after training (Helmreich, 1999). This approach has been typical of human factors programs in other industries as well. For example, air medical programs have been evaluated by questionnaires, which aimed to measure individuals’ attitudes towards team awareness and effective communication (Fisher et al., 2000). Similarly, crisis resource management training in medical operating rooms has been assessed by questionnaires, where participants where

---

1 In United Airlines Flight 811, a cargo door blew out on a Boeing 747 causing significant damage and the loss of two engines. United Airlines Flight 232, a DC10, suffered a catastrophic failure of the centre engine, resulting in the loss of all hydraulics and flight control systems.
asked to rate themselves on their performance (Holzman, Cooper, Gaba, Philip, Small & Feinstein, 1995).

This reliance on attitude measurement does not provide sufficient evidence for the effectiveness of human factors training. Surveying attitudes is relevant to some parts of program evaluation, such as judging employee reactions and detecting potential morale problems. However, when evaluating the benefits of a training program, attitudes are an unsuitable measure, for two reasons. Firstly, it is a debatable assumption that attitudes have a direct effect on behaviour. It is possible that trainees have simply learned the right thing to say and are responding appropriately. Even if trainees have absorbed the concepts being taught, attitudes give no indication of whether they know how to apply what they have learned. Secondly, and more importantly, attitudinal assessment is not appropriate for an objective evaluation of what an individual has learned. When psychologists wish to test a person’s intelligence, they do not give them a questionnaire asking them how intelligent they are. Instead, they use a set of performance items, which are compared to objective and standardised normative data points. Such behavioural methods of evaluation are applicable for human factors training programs.

Despite such flaws, many human factors training programs continue to rely on attitudes as a prime measure of program effectiveness. However, some sections of the aviation industry are shifting to behavioural methods of training and evaluation (see Flin & Martin, 2001 for a review). This illustrates another significant problem; namely that human factors training programs have not developed as a unified field with standardised methods. Rather, the area can be characterised as a diverse range of programs and concepts developed by separate airlines and government agencies (Salas et al., 1999b).
As a result, any evaluation of the effectiveness of human factors training needs to consider the diversity in the field, including the impact of different organisational cultures, the operating environments and the types of program conducted. Studies that claim to demonstrate the apparent failure or success of human factors training often need to distinguish which form or type of program is being used. All forms of human factors training are not created equal and without further probing of the type of courses participants underwent, a well-intentioned evaluation may not provide substantive evidence against the field as a whole.

**The use of behaviour based evaluation criteria**

In recent years, some human factors training programs have started to develop more rigorous criteria for training and evaluation. Behavioural competencies have begun to replace training for attitudes and awareness (eg. van Amermaete & Kruijisen, 1998). The development of behavioural competencies allows trainees to learn specific skills, which can be employed in work settings. Furthermore, the evaluation of specific competencies allows greater accuracy in assessing those areas in which individual trainees perform well, and those areas they need to improve. While behavioural human factors training is not universally applied, its acceptance appears to be spreading.

In contrast to the initial development of CRM style programs, characterised by their wide variety, different assumptions, different training methods, and a lack of common content, the recent growth of behavioural human factors training has been accompanied by a shift toward standardised programs. The involvement of regulatory bodies such as the Joint Aviation Authority (JAA) has led to increased standardisation of training methods and course content.
In the United States, the Federal Aviation Administration (FAA) has sponsored the development of Advanced Crew Resource Management (ACRM). This system is not a single package for all airlines, but a basis for individual airlines do develop their own programs. Significantly, the ACRM program includes skill-based, behavioural measures of crew performance, as well as standardisation of trainers.

In Europe, the Joint Aviation Authority has produced the NOTECHS system, an amalgamation of existing methods to measure non-technical skills. In this system, non-technical skills are defined as “attitudes and behaviours in the cockpit not directly related to aircraft control, system management, or standard operating procedures” (van Avermaete & Kruijisen, 1998). Non-technical skills are divided into four categories: cooperation, leadership and/or management skills, situation awareness, and decision making. The NOTECHS system also includes five principles, which are intended to provide objective assessment. The first requirement is that only observable behaviour is assessed. Secondly, for behaviour to be rated unacceptable, it is a requirement that there be a threat to flight safety. The third requirement is that unacceptable behaviour must be repeated during a check to determine if there is a substantive problem. Fourthly, each behaviour must be rated as either acceptable or unacceptable. Finally, an explanation is required for each unacceptable rating.

While the shift from attitude based evaluation to a focus on behavioural training and assessment of flight crew skills is promising (Flin & Martin, 2001), human factors training is now being increasingly applied to other areas of aviation such as maintenance (eg. Johnson, 2000b; Stelly & Poehlmann, 2000). With the broader application of human factors programs, it has become even more imperative to evaluate the evidence for their effectiveness.
Evidence for the effectiveness of human factors training

When considering the effectiveness of human factors training, it should be noted that some researchers believed that training effectiveness had already been validated before the development of behavioural programs like NOTECHS (Helmreich, 1996). However, most of the evidence cited above measured attitudes rather than behaviour, so that even when studies found that flight crew attitudes were improved (e.g., Helmreich & Wilhelm, 1991), this did not constitute reliable evidence. Indeed, some researchers concluded that evidence for human factors training success was sorely lacking (e.g., Novick, 1997; Wise, 1996).

With the spread of behavioural programs, there is now new scope for evaluating human factors training effectiveness. However, as behavioural programs have only been applied since the mid-1990s, empirical evaluations have been understandably limited in both military and commercial aviation contexts.

Human factors training effectiveness within Military Aviation

To date, the most convincing evidence of human factors training effectiveness within the military aviation environment has been published by Salas, et al., (1999a). In a recent paper, these authors reported two evaluation studies of CRM style training applied in naval aviation. The first study included 35 pilots and 34 enlisted aircrew from U.S. naval transport helicopters, taking part in the research program through their annual CRM training. The participants were trained in four skill areas (communication, assertiveness, mission analysis, and situational awareness) as part of a behavioural-focused program. The study used a
combination of attitudinal and behavioural indices, which were measured pre and post training. However, no control group was used. The results of the attitudinal evaluation were generally positive, which was consistent with previous research. More importantly, the study included a behavioural evaluation where participants received specific opportunities to demonstrate team behaviour. Trained teams performed significantly better than baseline teams on the teamwork behaviours. These results suggest that the CRM training produced at least short-term improvements in teamwork behaviour.

However, Salas et al., (1999a) noted one substantial problem with their first study. That is, organisational constraints had prevented them from establishing whether trainees were equally capable in team behaviour before the training commenced. Therefore, they conducted a second study with a similar structure to the first, but which incorporated a control group that received no training. This second study used 27 naval aviators, also taking part in the research as part of their annual CRM training. The training course addressed a greater variety of skill areas (decision making, assertiveness, mission analysis, communication, coordination, leadership, adaptability/flexibility, and situational awareness). The evaluation used similar attitudinal and behavioural indices to the first study, but the behavioural evaluation included a larger number of test items. This study found that teamwork behaviours improved when compared to the control group. Therefore, both studies reported by Salas et al., (1999a) demonstrated that the CRM training produced behavioural improvements in human performance.

The studies cited above are consistent with previous research by Salas and associates, which, demonstrate modest but consistent improvement in teamwork behaviour after CRM training, even with already experienced aviators. Further supporting evidence has come from other
areas of military aviation. In a comprehensive paper, Diehl (1991) reviewed the evidence for CRM within the military field with particular application to rotary wing aircraft, which tend to have higher accident rates compared to their fixed wing counterparts. Data from a variety of sources indicated that training programs reduced error rates. In one example, Bell Helicopters Textron Inc incorporated CRM training principles into crew operating their Jet Ranger helicopter. The global human error accident rate for this aircraft type fell 36% for the four-year period after the introduction of training, when compared to the preceding four-year period. In another instance, Petroleum Helicopter Inc, then the largest U.S. operator of commercial helicopters, introduced CRM training. The accident rate dropped 54% in the following two years, whereas it had remained relatively stable in the six preceding years. However, the most substantive evidence came from the USAF Military Airlift Command (MAC). In 1985, MAC introduced a CRM-style program. According to Diehl (1991), in the following five-year period, accident rates fell 52% and serious flight-related mishaps fell 51% when compared to the preceding five-year period. Accident and serious mishap rates in other parts of the USAF fell 18% and 21% respectively in the corresponding periods. The main difference between MAC and other air commands was CRM training.

While the above research does seem to support the effectiveness of human factors training, it has one substantial drawback: the evidence comes from military applications. The military is different from commercial aviation in both its operating environment and its training programs. For example, military aviation accepts relatively risky procedures, which would not be tolerated in the conservative commercial arena. This means that the impact on teamwork in training will also be different, both in the outlook of pilots and in the specific behaviours used as skill markers. Nevertheless, the methodology developed for evaluation within the military appears to be suitable for adaptation in a commercial environment.
Behavioural evidence for the effectiveness of human factors training within commercial aviation has been scarce to date. Historically research based on positive attitude change has provided evidence supporting the impact of training programs. For example, Byrnes and Black (1993) utilised the Cockpit Management Attitudes Questionnaire (CMAQ; Gregorich & Wilhelm, 1993) to assess attitude change in a group of U.S airline pilots, and found that human factors training produced a positive attitude shift, which remained stable for up to five years. Other studies using the CMAQ have supported these results reporting a positive shift in pilot attitudes towards human factors concepts following a training course (Gregorich, 1993; Gregorich, Helmreich & Wilhelm, 1990; Incalcaterra & Holt, 1999).

In perhaps the best empirical evaluation before the advent of behavioural human factors training, Helmreich & Foushee (1993) reported on a program, which included the use of line-oriented flight training (LOFT) and periodic training. They found that ratings of human performance in line operations improved substantially after training. While this study did not use specific behavioural markers it does provide additional supporting evidence for the use of human factors training within commercial operations.

More recent attempts to evaluate human factors training effectiveness within aviation have focused on behavioural change. Boehm-Davis, Holt & Seamster (2001) report on a training program, which used Line Orientated Evaluation’s (LOE’s) and line checks as performance indicators. Data from the first year, before the introduction of the program, functioned as a baseline, and was compared with data from the next two years. The participants were evaluated using standardised behavioural assessments. From this, the authors deduced that the
training improved observable performance in CRM skills. The authors argued that this study provided supporting evidence to a previous application of the LOE in which one fleet (human factors training) outperformed another (non-human factors training) on half of the check items (Holt, Boehm-Davis & Hansberger, 1999).

Other studies reporting behavioural change following human factors training include Clothier (1991) who used the LINE LOS Checklist (LLC). Comparison between trained and untrained crews in LOFT and line observations indicated that there was a significant difference after training on 12 of the 14 LLC categories. This was supported by Helmreich and Foushee (1993) who found significant positive differences on all 14 categories of the LLC, over a three year period, following human factors training.

Gunther (2000) reported that over a two year period at Continental Airlines, the introduction of the Line Operations Safety Audit (LOSA), a non-jeopardy, normal operations audit program, and subsequent introduction of human factors training in threat and error management, resulted in some remarkable improvements on behavioural safety indices. In 1996, studies indicated that on average flight crew detected 15% of internal errors. After the introduction of a threat and error management course, 55% of internal errors were detected. In the same period, flight crews achieved a 78% reduction in unstable approaches at 1000ft above ground level, and a 40% reduction in unstable approaches at 500ft.

While the papers by Boehm-Davis et al., (2001), Clothier (1991), Gunther (2000) and Helmreich and Foushee (1993) are some of the few published studies that can be found in commercial aviation, there appears to be a lack of scientifically rigorous research, which in part has ensured that there remains little confidence in the effectiveness of human factor
training programs. The few studies that are published tend to rely on a combination of rational and anecdotal information, do not use control participants, and do not clearly specify the intervention that was applied. Furthermore, there does not appear to be any studies that have attempted to measure where human factors training programs demonstrate a return on investment. In contrast, it is the aviation maintenance area, which is leading the way in cost effectiveness research.

**Evidence for the cost effectiveness of human factors training**

Several studies have examined the financial efficacy of human factors training on aviation maintenance. These have generally found investment returns on training programs, in the form of reductions in equipment damage incidents, employee down time, or lost time injuries. These reductions have provided significant cost savings for operators.

Johnson (2000b) reported on the outcome of a two day human factors training course for maintenance staff, which led to a reduction of ground towing damage events by 75%, with a resultant cost saving of US$195,000 per year. Similarly, Eiff (2000) described the implementation of a human factors education program on shift turnover. This demonstrated a reduction in average man-hours (by 2474 hours) to complete a “D” check, with a reported cost saving of US$94,000 per year. In addition, Stelly & Poehlmann (2000) reported on the outcome of a two day human factors training course for Continental Airlines. The course produced several behavioural outcomes; a 68% reduction in ground damage incidents, a 12% decrease in job injuries, and 10% reduction in staff overtime. The total cost saving was US$60,000 per year for five years. However, the most impressive cost benefit research to date is that of Taylor (2000). The introduction of a maintenance resource management program at
a U.S. airline reduced lost-time injuries (LTI’s) by 80% over two years, with a total claimed cost saving of US$1,300,000 over that period.

While, the above maintenance studies provide useful cost efficacy data there are no published studies in other areas of aviation, such as ramp, freight or catering. Furthermore, several methodological weaknesses limit their application.

**Flaws with current human factors training evidence**

The most common problem of many evaluations of aviation human factors training is the lack of a control group to determine if any observed effects are a result of the training, and not some other cause. This is a shortcoming across other industries as well (eg. Fisher et al, 2000; Holzman et al, 1995). Of the studies reviewed in this paper, the vast majority did not include a control group to verify the effects of training.

A second methodological weakness is the relative lack of longitudinal evaluation. Short-term behavioural evaluation methods give a useful indication that human factors training may be effective, but without a longitudinal evaluation, it is impossible to determine if the effects of training diminish over time, and if so, does recurrent training have any effect? While the review by Diehl (1991) included some examples of longitudinal evaluation, these lacked formal control groups. Furthermore, it is unfortunate that an otherwise very impressive study by Salas et al., (1999a), lacked longitudinal evaluation. Some researchers have included longitudinal evaluation, and others have included control groups, but none have reported the application of both.
The third methodological weakness is specific to the Salas et al., (1999a) paper. While this paper provided useful behavioural data, the relatively low numbers of participants weakens what conclusions can be drawn. Given cost and logistical constraints, it is understandable why this and similar studies use a small number of participants, but the lack of large sample sizes nonetheless remains a problem when deciding what weight should be given to research findings.

A fourth methodological shortcoming is a lack of information about training program cost-effectiveness, with the exception of the maintenance studies cited above. While studies into aviation maintenance provide measurements of cost, savings and return on investment, similar information is unavailable for other areas such as flight operations or the airport/ramp environment. This is most probably due to the costs of incidents being harder to document outside of maintenance. What is lacking is a definable set of cost indices, which can be used to measure the effectiveness of a human factors training program. Moreover, the maintenance studies that have been cited in this paper do not appear to have been published in peer-reviewed literature. Therefore, it is difficult to determine if the claims made in return of investment are valid.

From a comprehensive review of 48 studies that claimed to evaluate the effectiveness of CRM training in the aviation industry, O’Connor, Flin and Fletcher (2001) suggest, that while there appears to be some evidence demonstrating a positive change in attitude, knowledge and behavior following CRM training, training evaluation is not the norm. This is despite the prolific use of human factors training programs throughout the aviation industry.
Conclusions and recommendations

For those aviation organisations contemplating human factors training, or who have an existing program but wish to expand it into other areas such as maintenance or airport/ramp operations, do the studies reviewed above provide conclusive evidence for justifying capital expenditure?

The return on investment studies (Stelly & Poehlmann, 2000; Taylor, 2000) conducted within aviation maintenance provide some solid evidence upon which to base a business case. However, the cost effectiveness evidence from operational areas outside of aviation maintenance, are lacking despite some recent studies in the flight operations environment supporting behavioural safety improvements following the introduction of human factors training (Boehm-Davis et al, 200; Gunther, 2000).

Clearly, more cost effectiveness data is needed to add weight to the argument. Measuring costs is relatively easy in some areas of aviation, as the evidence has shown in maintenance. Other areas, such as airport/ramp operations, also have relatively quantifiable measures of cost efficacy, although the evidence for the effectiveness of human factors training to date is less clear-cut in this area (eg. McDonald, Cromie & Ward, 1997).

One of the most promising indicators appears to be lost time injuries (LTI’s). LTI’s represent direct, measurable costs to airlines. They also provide an objective measurement, which can be used as a criterion for evaluating human factors training. It has been estimated that 96% of LTI’s are behavioural in nature (Dupont, 2000). Given that the focus of contemporary human
factors training programs is to modify and reinforce effective safety behaviour, it is reasonable to assume that LTI’s will be reduced. A reduction in LTI’s from human factors training has already been demonstrated in some aviation maintenance human factors programs (eg. Taylor, 2000). One current study being conducted at Qantas Airways (Edkins, 2000; Edkins & Pietrovitch, in preparation) may provide further supporting evidence. This longitudinal Australian study is being conducted within airport/ramp operations to determine the impact of a behaviour based human factors training program on a number of safety indices including workers compensation claims (LTI's).

Those researchers intending to evaluate the effectiveness of human factors training programs need to ensure that an appropriate methodology is employed, that includes sound evaluation criteria. The training should be evaluated by behavioural-based methods, using specific competencies to assess non-technical skills. Such evaluations should be longitudinal, otherwise any observed effects may merely be manifested in short-term benefits. Control groups are also needed to determine if training has produced any observed effects. O’Conner et al (2001) suggest a multifaceted evaluation approach that employs a combination of attitude, knowledge, behavioural and organizational measures.

Furthermore, human factors programs implemented across the aviation industry need to more consistently report on cost indices, as a measure of program success. These indices may include objective criteria such as reduction in LTI’s, equipment damage, employee work time, employee over time etc. Tracking and reporting on costs will help to ensure that operators, who are struggling to rationalise expenditure, have a stronger business case upon which to justify an expected return on investment.
References


Du Pont (2000). *Briefing to Qantas safety management by David Hainsworth, Senior Account Manager*. 26 October, Melbourne.


