Running Head: TAXONOMIES, REPORTING SYSTEMS, AND DATA COLLECTION TOOLS

A Review of Selected Aviation Human Factors Taxonomies, Accident/Incident Reporting Systems, and Data Collection Tools

J. Matthew Beaubien

David P. Baker

American Institutes for Research

1000 Thomas Jefferson Street, NW

Washington, DC 20007

October 24, 2002

Beaubien, J. M., & Baker, D.P. (2002). A review of selected aviation Human Factors taxonomies, accident/incident reporting systems, and data reporting tools. *International Journal of Applied Aviation Studies, 2*(2), 11-36.

Abstract

The recent introduction of the Aviation Safety Action Program (ASAP) has provided carriers with a powerful tool for understanding the Human Factors issues that their crews face during typical line operations. Unfortunately, this tool has rarely been used to its fullest potential. We believe that this is because most ASAPs collect Human Factors data using text narratives that require costly and time-consuming content analysis. To address this problem, we are developing a generalized Human Factors taxonomy and electronic data reporting/analysis tool that will be provided at no cost to industry. This review – which summarizes aviation error reporting, classification, and analysis systems from the United States, Great Britain, and Australia – serves as the foundation for our future taxonomic research.

A Review of Selected Aviation Human Factors Taxonomies,

Accident/Incident Reporting Systems, and Data Collection Tools

The Aviation Safety Action Program (ASAP), which collects de-identified incident reports from line pilots regarding threats to safety (Federal Aviation Administration, 2000), is very similar to NASA's Aviation Safety Reporting System (ASRS; Federal Aviation Administration, 1987). For example, both systems encourage pilots to self-report their errors by providing limited immunity from prosecution. However, unlike the ASRS, the ASAP is carrier-specific. Drawing on the ASRS example, many ASAPs collect background information regarding the pilot who submits the report and the flight conditions that immediately preceded the event. In addition, space is usually provided so that the reporting pilot can write a short narrative describing the event, the causal factors that precipitated it, and suggestions for preventing its reoccurrence. As a general rule, most of the Human Factors issues described in ASAP reports can be found within these narratives.

In a perfect world, voluntarily-collected safety data such as ASAP incident reports, line audit summaries, and Flight Operations Quality Assurance (FOQA) output would be used to continuously assess carrier safety and to suggest recommendations for improvement. Unlike line audits and FOQA output that are designed to answer the question "What happened?", ASAP reports are uniquely designed to answer the question "Why did this event happen?". Unfortunately, because the analysis of textual data is extremely difficult, ASAP data has rarely been used to its fullest capacity, such as in developing training objectives, Line Operational Simulation (LOS) scenarios, and other safety-related interventions (for a notable exception, see Baker, Chidester, & Brannick, 1998). To remedy this problem, we are developing a generalized

Human Factors taxonomy and electronic reporting/analysis tool that will be made freely available to industry. This tool will use pull-down menus and check-in-the-box items to supplement the text narrative, thereby allowing carriers to easily classify and quantify the root causes of pilot error. Eventually, this electronic reporting tool will include sophisticated data analysis and graphing tools to help carriers identify the most pressing problems, develop data-driven interventions, and assess the effectiveness of these interventions by statistically comparing pre- and post-intervention data. This review – which summarizes the state-of-the art in aviation error reporting, classification, and analysis – serves as the foundation for our future taxonomic research.

Method

Literature Review

We began our review by searching the PsycINFO database, the Dissertation Abstracts International database, and recent International Symposium on Aviation Psychology conference proceedings for published studies containing the following key words: "Aviation Safety Action Program," "ASAP," "Aviation Safety Reporting System," "ASRS," "incident reporting system," "accident reporting system," "human factors taxonomy," and "human error taxonomy." Next, we expanded our search to include unpublished documents on the Internet using the same key words. We then scanned each document's reference list for additional source materials. Finally, we contacted the authors and administrators of these systems for unpublished technical reports, in-press manuscripts, and briefing slides.

Criteria for Evaluating Taxonomies of Human Performance

Few methods have been proposed for evaluating the effectiveness of classificatory systems. Much of the work in this area was conducted by Fleishman and his colleagues

(Fleishman & Quaintance, 1984; Fleishman & Mumford, 1991), who identified three major criteria: internal validity, external validity, and utilitarian criteria.

Internal Validity. Internal validity concerns the extent to which a classification system is logically organized and parsimonious (Fleishman & Quaintance, 1984). There are five issues to consider when assessing the internal validity of a taxonomy. The first issue involves the reliability of the descriptors that are used to classify an event. All things being equal, taxonomies that can be reliably used by novices are more internally valid than taxonomies that require special training. The second issue involves the reliability of the classification system as a whole. All things being equal, taxonomies that allow an event to be reliably coded despite random fluctuations in the sampling of descriptors (i.e., the textual information contained in the ASAP narrative) are more internally valid than taxonomies that do not.

The third and fourth issues, respectively, involve the taxonomy's use of mutually exclusive and exhaustive descriptors. All things being equal, taxonomies that use mutually exclusive and exhaustive descriptors are more internally valid than taxonomies that do not. The fifth and final issue involves the statistical clustering of variables. All things being equal, taxonomies that reveal meaningful and interpretable patterns, for example via factor analysis or cluster analysis, are more internally valid than taxonomies that do not. In summary, Fleishman's research suggests that the most internally valid taxonomies are those that can be reliably used by novices, that can reliably categorize events despite random fluctuations in the wording of the narrative text, that use mutually exclusive and exhaustive descriptors, and that reveal meaningful patterns.

External Validity. External validity concerns the extent to which a taxonomy achieves the objectives for which it was designed (Fleishman & Quaintance, 1984). There are three major issues to consider when evaluating the external validity of a classification system. The first issue concerns the generalizability or robustness of the taxonomy's findings. All things being equal, taxonomies that can classify new cases with the same degree of accuracy as the "test cases" that were used to develop them are more externally valid than those that do not. The second issue concerns the taxonomy's capacity to help researchers plan programmatic lines of research. All things being equal, taxonomies that provide direction for future research by differentiating among "what we know," "what we don't know," and "what we need to know" are more externally valid than taxonomies that only summarize what is known. The third and final issue concerns the taxonomy's predictive validity. All things being equal, taxonomies that help researchers accurately predict behavioral outcomes given a known set of initial predictors are more externally valid than taxonomies that do not permit prediction. In summary, Fleishman's research suggests that the most externally valid taxonomies are those that cross-validate with new data sets, that identify gaps in the available research, and that predict meaningful outcomes.

Utilitarian Criteria. Utilitarian criteria concern the extent to which a taxonomy is useful and efficient (Fleishman & Quaintance, 1984). There are four major issues to consider when evaluating the utility of a classification system. The first issue concerns the extent to which a taxonomy promotes communication among different user groups. All things being equal, taxonomies that promote communication among users are more useful than those that do not. The second issue concerns the extent to which a taxonomy is used to solve applied problems. All things being equal, taxonomies that are used to solve applied problems are more useful than those that are developed solely for research purposes. The third issue concerns the extent to which the user must expend resources such as time and training to use a taxonomy effectively. All things being equal, taxonomies that require fewer start-up costs are more useful than those that require a greater investment of resources. The fourth and final issue concerns the number of people and organizations who use a given taxonomy on a regular basis. All things being equal, taxonomies with a large user base are more useful than those with a smaller user base. In summary, Fleishman's research suggests that the most useful taxonomies are those that facilitate communication by standardizing the terminology within a domain, that are used to solve applied problems, that require few start-up costs, and that have a large user base.

Proposed Evaluation Strategy. Unfortunately, many of the systems that we identified were poorly documented, thereby making it difficult to apply Fleishman and colleagues' criteria in their original form. Even after contacting the systems' developers, many questions remained. Therefore, we compared and contrasted each system along six dimensions: mission statement, theoretical orientation, data organization and management, validation evidence, uses/utility, and strengths and weaknesses. These dimensions were chosen because they apply equally well to describing taxonomies of human performance, accident/incident reporting systems, and data collection tools. They also allowed us to compare these different systems using a common set of metrics.

Fortunately, our evaluation dimensions can be roughly classified within Fleishman and colleagues' framework. For example, because our dimensions of "mission statement" and "theoretical orientation" focus on how the data are collected and organized, they address the issue of internal validity. Similarly, because our dimension "validation evidence" focuses on how the systems are used, it addresses the issue of external validity. Finally, our dimension "uses/utility" addresses both external validity and utilitarian issues. Nevertheless, there are certain differences between Fleishman's criteria and our evaluation dimensions. For example, whereas several of Fleishman's criteria (e.g., inter-rater reliability) can be assessed quantitatively, most of our dimensions (e.g., "theoretical orientation") can only be assessed qualitatively. Despite these differences, we have attempted to apply Fleishman and colleagues' criteria as faithfully as possible.

In the following sections, we compare and contrast several widely used accident/incident reporting systems, aviation Human Factors taxonomies, and data collection tools. Our findings are summarized graphically in Tables 1-3. We caution the reader that our review is selective, focusing only on those systems that have direct relevance for building a generalized Human Factors taxonomy for use in ASAP. For a more comprehensive review of the methods and tools used in flight safety data analysis, the reader is directed to works by Ingebretson and Morrison (1999) and the Global Aviation Information Network (2001).

Accident/Incident Reporting Systems

In this section, we describe some of the earliest accident/incident reporting systems, such as NASA's Aviation Safety Reporting System (ASRS) and its counterparts in the United Kingdom. These systems were developed with the extremely broad mission of improving all aspects of their home countries' national aerospace systems. As a result, they are best suited for identifying broad safety trends. These systems are somewhat less useful for identifying the specific problems faced by a particular carrier, or for suggesting targeted interventions to address those problems.

Aviation Safety Reporting System (ASRS)

Mission Statement. The Aviation Safety Reporting System (ASRS) is a voluntary, confidential, and non-punitive incident reporting system that was designed to promote aviation safety by encouraging pilots to report unsafe occurrences and hazardous situations (http://asrs.arc.nasa.gov). Although funded by the FAA, the program is maintained by NASA, which collects, de-identifies, catalogs, and analyzes the incident reports. NASA's participation has helped to create a safety culture that encourages pilots to openly discuss and learn from their errors (National Aeronautics and Space Administration, 1999).

The goals of ASRS are to improve the current national aviation system and to provide data for future system planning (Federal Aviation Administration, 1987). ASRS reports may be submitted by pilots, air traffic controllers, cabin crews, and mechanics. Feedback – in the form of periodic bulletins and newsletters, database search requests, quick-turnaround data analyses, and other products – is provided to air carriers, researchers, and government agencies so that they can investigate allegations of unsafe practices and take corrective action (National Aeronautics and Space Administration, 1999).

Theoretical Orientation. There is little publicly available information concerning how ASRS was developed. Based on our review, it does not appear that the ASRS system or its associated reporting forms were developed according to any particular theory of human error. The data fields were culled from those commonly used to describe previous accidents and incidents, and which could be easily quantified or classified into discrete categories.

Data Organization and Management. ASRS provides separate reporting forms for pilots, air traffic controllers, cabin crews, and mechanics. Each form contains a series of fields that describe the relevant conditions (e.g., type of flight, phase of flight, weather) that immediately preceded the incident. Space is also provided for a text narrative so that the reporter can provide additional details, such as causal/contributing factors, the chain of events, and suggestions for preventing the event's reoccurrence. As mentioned earlier, there is little information regarding how ASRS was developed and how it is currently maintained. As a result, conclusions regarding how ASRS reports are stored electronically and analyzed remain speculative. Our review of publicly available ASRS report sets (http://asrs.arc.nasa.gov/report_sets.htm) suggests that many of the descriptive fields are stored as nominal variables, possibly with pre-defined response options. The incident narratives are stored as text.

Validation Evidence. All ASRS reports are independently coded by at least two analysts, who are recruited from the ranks of retired airline pilots, air traffic controllers, and Human Factors researchers (National Aeronautics and Space Administration, 1999). The analysts have two major tasks. Their first task is to diagnose the underlying cause(s) in each incident report. Their second task is to identify hazards that require immediate attention, and to forward this information to the appropriate regulatory authority. To date, we have been unable to identify any data regarding the average level of agreement between pairs of analysts. We have also been unable to identify any data regarding the accuracy or comprehensiveness of their ratings.

Uses/Utility. In addition to the information products that are developed by ASRS staff, the full-text of all ASRS incident reports are available to the public for secondary analysis (http://nasdac.faa.gov/asp/asy_asrs.asp). In recent years, this wealth of data has been mined to better understand the nature and frequency of pilot errors (Sarter & Alexander, 2000), factors that inhibit the development of situational awareness (Jentsch, Barnett, Bowers, & Salas, 1999), and the effects of time pressure on pilot decision-making (Stone, Babcock, & Edmunds, 1985).

Strengths and Weaknesses. A major strength of ASRS is that the pilot reporting form contains a section for describing a second aircraft (if relevant) that was involved in the event. A second strength is that ASRS explicitly inquires about the chain of events. Specifically, the reporter is asked to describe how the problem arose, what factors contributed to the event, and what corrective actions were taken. A third strength is that ASRS inquires about errors of omission as well as commission.

Despite these strengths, our research suggests that virtually all of the Human Factors-related information in an ASRS report is stored as text. A more convenient and effective way of collecting such information might be to include Human Factors “reason codes” which the reporter could select via menus or check-in-the-box items. This information could supplement the information provided in the narrative. Not only would this help standardize the data collection process, but it could also serve as a validity check for assessing the analysts’ diagnosis of the event.

Confidential Human Factors Incident Reporting Programme (CHIRP)

Mission Statement. The UK Confidential Human Factors Incident Reporting Programme (CHIRP) is a voluntary, confidential, and non-punitive incident reporting system that was established to identify and resolve a wide range of safety-related issues in the UK air transport industry. CHIRP is maintained by the CHIRP Charitable Trust, an independent organization that is funded by the UK Civil Aviation Authority (CAA) and the UK air transport industry (http://www.chirp.co.uk).

Incident reports may be submitted by general aviation pilots, commercial airline pilots, cabin crews, air traffic controllers, engineering and maintenance staff, and design and production staff. These reports are de-identified by CHIRP staff and analyzed for trends. Feedback – in the form of periodic newsletters – is provided to each constituency group. Each issue of their FEEDBACK newsletter focuses on a subset of “high priority” issues that have been identified by CHIRP analysts. Each newsletter features summary statistics, de-identified incident reports on selected topics of interest, suggestions for avoiding reoccurrences, and references to relevant CAA regulations.

Theoretical Orientation. Like ASRS, there is little publicly available information regarding how CHIRP was developed and how it is currently maintained. Based on our review, it does not appear that the CHIRP system and its associated incident reporting forms were developed based on any particular theory of human performance. Based on our review, it appears that CHIRP was modeled directly on the ASRS system.

Data Organization and Management. CHIRP provides specific reporting forms for each constituency group. Each form contains a series of fields that describe the relevant conditions (e.g., type of flight, phase of flight, weather) that immediately preceded the incident. Space is also provided for a text narrative. Like ASRS, many of the descriptive fields are stored as nominal variables and that the incident narratives are stored as text. Finally, we speculate that CHIRP's glossary of aviation-related terms (http://www.chirp.co.uk) is used, in part, to classify Human Factors issues in these narratives.

Validation Evidence. We have been unable to identify any data concerning the reliability or validity of CHIRP's classification system. Because CHIRP is not a Human Factors taxonomy *per se,* it is not unexpected that validation studies would be difficult to locate. The only evidence that we were able to locate is a user satisfaction survey that contained over 28,000 responses from line pilots, flight engineers, and air traffic controllers. An overwhelming majority of the respondents indicated that CHIRP makes a useful contribution to safety (99.0%), provides useful Human Factors information (96.4%), and provides an independent approach to reviewing reports (96.8%) (CHIRP, 1999). Based on the results, UK pilots clearly like the CHIRP program. However, given the dearth of publicly available information about CHIRP, it is unlikely that pilots have a detailed understanding of how the program actually operates.

Uses/Utility. Unlike ASRS, the CHIRP database is not made available to the public for secondary analysis. We have also been unable to locate any published or unpublished studies that are based on CHIRP data. However, this may be due to the fact that certain search engines (e.g., PsycINFO) exclude European journals.

Strengths and Weaknesses. A major strength of CHIRP is its inclusion of a comprehensive glossary. While it is not known to what extent this glossary is used in classifying CHIRP incident reports, a standardized glossary is essential when coding textual data, especially when multiple coders are used (Creswell, 1994). Another strength is that like ASRS, CHIRP does not limit itself to collecting incident reports from only one constituency group. In theory, this could allow CHIRP staff to identify mutli-faceted perspectives to individual incidents by linking pilots' reports with ATC reports for the same incident. This could be achieved, for example, by sorting reports based on their date, time, and location of occurrence. A third strength is that CHIRP frequently provides constituency-specific feedback regarding common Human Factors problems and suggested solutions.

Despite these accolades, CHIRP has its limitations. Like ASRS, virtually all of the Human Factors-related data is located exclusively in the text narrative. Another limitation is that CHIRP does not provide a formal mechanism for specifying the chain of events. Finally, CHIRP does not formally request information on contributing/mitigating factors, how the problem was discovered, or suggested corrective actions.

Confidential Aviation Incident Reporting (CAIR)

Mission Statement. The Confidential Aviation Incident Reporting (CAIR) system is Australia's analog to ASRS and CHIRP. Like its counterparts, CAIR is a voluntary, confidential, and non-punitive incident reporting system that was established to proactively identify safety-related deficiencies and to suggest appropriate remedies. However, CAIR's focus is not on individual events, but on systems, procedures, and equipment (http://www.atsb.gov.au/aviation/cair/index.cfm). CAIR is one of several incident reporting systems and risk management tools developed by the Bureau of Air Safety Investigation (BASI), a quasi-independent branch of the Australian Civil Aviation Safety Authority (CASA) (Lee, 2001).

The CAIR system is open to anyone who wishes to submit a report, including pilots, cabin crews, air traffic controllers, maintenance, and passengers (http://www.atsb.gov.au/atsb/facts/cair_program.cfm). All reports are de-identified by CAIR staff and then analyzed for systems-level issues. Feedback is provided in a supplement to CASA's Flight Safety Australia magazine and BASI's Asia-Pacific AIR SAFETY magazine. These magazine articles are designed to disseminate information about BASI initiatives, to identify relevant safety issues, and to educate the aviation community. The magazine articles use a similar format to those used by ASRS and CHIRP.

Theoretical Orientation. Like ASRS and CHIRP, there is little publicly available information regarding how the CAIR program was developed and how it is currently maintained. Based on our review, it does not appear that CAIR and its associated incident reporting form were developed based on any particular theory of human performance. It appears that CAIR was modeled directly on the ASRS system.

However, BASI takes a unique position to analyzing CAIR data. Specifically, BASI analysts never use the terms "primary cause" or "contributing cause." Rather, all results are presented as "findings" and "significant factors" (Lee, 2001). This is because BASI analysts consider all air safety occurrences to be a complex interaction of many factors. Similarly, BASI does not distinguish between accidents and incidents. They are all classified as "safety occurrences" (Lee, 2001). These two issues – not assigning causality to a single factor, and not distinguishing between accidents and incidents – also appear in Reason's (1990) treatment of human error. Even though the CAIR incident reporting form may not have been designed according to any particular theory of human performance, the data collected via CAIR appear to be analyzed and interpreted consistent with the Reason (1990) model.

Data Organization and Management. CAIR uses a generic reporting form that can be completed by all respondents. The form contains a series of fields that describe the conditions (e.g., type of flight, phase of flight, weather) that immediately preceded the incident. Space is also provided for a text narrative. In many ways, the CAIR incident reporting system is identical to ASRS and CHIRP. Many of the descriptive fields are stored as nominal variables, possibly with pre-defined response options, and the incident narratives are stored as text.

Validation Evidence. To date, we have been unable to positively determine what type of taxonomy that CAIR uses for classifying and analyzing Human Factors issues. Given CAIR's similarities with ASRS and CHIRP, we speculate that they use a similar taxonomic structure. Like ASRS and CHIRP, it appears that these taxonomies were developed in an ad hoc manner.

Uses/Utility. CAIR data has reportedly been used to identify safety deficiencies in various areas, including flight operations, passenger handling, cargo handling, maintenance, air traffic control, and airport security. However, we have been unable to locate any published or unpublished studies that are based on CAIR data. Such reports may be available, perhaps in limited circulation.

Strengths and Weaknesses. A major strength of the CAIR system is its holistic approach to air safety investigations. By not distinguishing between "primary" and "contributing" causes, BASI analysts are better able to engage in systems-level thinking. At the same time, by not making such distinctions, all causal factors – even the most trivial ones – may artificially receive equal weight. As a result, this practice may both hinder and help the search for causal trends.

Taxonomies of Human Performance and Human Error

In this section, we describe several taxonomies of human performance and human error, such as the Human Factors Analysis and Classification System (HFACS). These taxonomies were specifically designed to classify Human Factors issues in accident/incident reports and line audits. As a result, they are somewhat better suited to identifying the specific problems faced by an individual carrier and for suggesting targeted interventions to those problems.

Human Factors Analysis and Classification System (HFACS)

Mission Statement. The Human Factors Analysis and Classification System (HFACS) is a general taxonomy of human error that was designed to provide a comprehensive framework for identifying accidents' causal factors, developing data-driven interventions, and objectively evaluating the effectiveness of these interventions (Shappell & Wiegmann, 2000). The HFACS taxonomy takes a multivariate approach to accident analysis by explicitly modeling active and latent threats to safety.

Theoretical Orientation. The HFACS taxonomy is based on Reason's (1990) model of human error. At its most general level, HFACS distinguishes between active and latent threats to safety. Active threats refer to unsafe acts (i.e., errors and violations) that are committed by the crew immediately preceding the accident or incident. Although highly visible, unsafe acts are extremely difficult to predict or control (Reason, 1990; Shappell & Wiegmann, 2000). More perilous are latent threats that often lie undetected for extensive periods of time. Latent threats refer to preconditions for unsafe acts, unsafe supervision, and organizational influences that compromise safety by allowing unsafe acts to proceed unchecked or which intensify their effect. Although latent threats can be difficult to detect and expensive to remedy, they have great potential for reducing the number of accidents.

Data Organization and Management. HFACS analyses begin by first identifying relevant accidents and then collecting their associated National Transportation Safety Board (NTSB) or Department of Defense (DoD) accident summary reports. Each accident report typically includes a wealth of information, including the official cause(s) of the accident, characteristics of the aircrew, relevant events that occurred up to 48 hours prior to the accident, and summaries from the aircraft's maintenance log (U. S. Navy, 2001). Next, the official causal factors are identified from each report. These causal factors are then coded by identifying the corresponding field in the HFACS taxonomy. Finally, the results are summarized across the entire sample of accidents. Additional descriptive information is often reported in HFACS studies. This may include accident locations, meteorological conditions, or other relevant factors. However, because these fields are not part of the HFACS taxonomy proper, they are not consistently reported from study to study.

Validation Evidence. To date, all HFACS validation research has been conducted by the system's developers using post-accident summary reports. Despite the fact that multiple researchers code each accident report, measures of inter-rater agreement are not consistently reported from study to study. When they are reported, however, they tend to be favorable. For example, an analysis of 119 aircrew-related accidents (Wiegmann & Shappell, 2001) revealed a kappa value of .71, which is considered "good" (Fleiss, 1981). However, because the researchers are also the taxonomy's developers, it is unclear to what extent these results will generalize to analysts who have far less training using HFACS.

Content validation studies using HFACS have been few, but positive. In one study (Wiegmann & Shappell, 2001), the HFACS taxonomy accommodated all 319 human causal factors that were identified in the sample of NTSB accident reports. All but two of the HFACS categories were used at least once, and no additional category codes had to be created. However, supervisory and organizational factors have been rarely identified as contributing causes of aviation-related mishaps (Pape, Wiegmann, & Shappell, 2001; Wiegmann & Shappell, 2001). These results may reflect the difficulties associated with collecting such information.

Uses/Utility. The HFACS taxonomy has been applied to the analysis of aircrew-related (Wiegmann & Shappell, 2001) and ATC-related (Pape, Wiegmann, & Shappell, 2001) accidents in military and civilian aviation. However, we have been unable to identify any studies where HFACS has been used to analyze incident reports, which generally contain much less information than official accident reports.

Strengths and Weaknesses. HFACS has a number of desirable qualities. First, it is based on a sound theoretical model of human error. As a result, it may be possible to adapt HFACS for use in a variety of ASAP programs, such as those involving maintenance or dispatch. Second, the taxonomy is organized in an efficient, hierarchical structure. This should reduce the cognitive demands on the user, for example if the HFACS taxonomy were to be embedded within an incident reporting form. Finally, content validation studies suggest that HFACS is relatively comprehensive, thereby requiring few additional fields.

Notwithstanding, HFACS does have its shortcomings. First, depending on the user's needs, the HFACS coding system may be too coarse to identify specific operational problems or to suggest interventions for those problems. For example, simply identifying a problem as resulting from "inadequate supervision" or "organizational climate" provides few clues for remedial action. Supplementing the HFACS taxonomy with "reason codes" might provide more detailed information for identifying specific areas of improvement. Second, even though latent threats are identified as important leverage points for facilitating change, it has proven exceedingly difficult to collect such information, even when using detailed accident reports. This problem would likely be exacerbated when analyzing the short narratives that are included in incident reports. Finally, the HFACS taxonomy does not identify the chain of events. As a result, it is difficult to separate causes from effects, even at the most granular level.

Approach-and-Landing Accident Coding Form

Mission Statement. The Approach-and-Landing Accident Coding Form is a data collection instrument that was developed by the Approach-and-Landing Accident Reduction (ALAR) task force. The task force had two primary goals. The first was to identify the causes of accidents that occur during the approach and landing phases of flight. The second was to develop practical recommendations for reducing the accident rate (Khatwa & Helmreich, 1998).

The coding form was developed to organize the data collected during the task force's review of several hundred accident reports and line audits. As the name implies, the coding form is not a taxonomy *per se*. However, it does contain several coding systems for classifying background information, the primary and contributing causes of accidents, and the consequences of those accidents.

Theoretical Orientation. The Coding Form takes an eclectic approach to collecting, organizing, and analyzing accident-related data. It draws on previous accident taxonomies (Khatwa & Roelen, 1997), Reason's (1990) model of human error, and the Line/LOS checklist (Helmreich, Butler, Taggart, & Wilhelm, 1995). The Coding Form is organized in an efficient layout. The first four pages are devoted to collecting demographic information concerning the flight, the flight crew, the external environment, and the airport/ATC/approach. The next three pages are devoted to collecting information regarding the accident's primary and contributing causes, the raters' confidence in his/her assessment of these causal factors, and the accident's consequences. The eighth page is devoted to assessing crewmembers' teamwork interactions. The ninth and final page is devoted to identifying accident prevention strategies.

Data Organization and Management. Information concerning the Coding Form and its accompanying database appears limited to that described in Khatwa and Helmreich (1998). As a result, it is difficult to draw firm conclusions regarding how the data were stored in an electronic database, coded, and analyzed. Most of the fields were stored as nominal variables, possibly with pre-defined response options, and that the accident prevention strategies were stored as text. Our review suggests that only one field, the primary and contributing causes, was organized in a truly hierarchical format.

Validation Evidence. The Khatwa and Helmreich report does not describe the procedure used to pilot test the Coding Form prior to full-scale data collection. Moreover, because group consensus judgments were used to generate the final ratings, measures of inter-rater agreement (e.g., collected prior to the consensus meetings) were not available. Finally, with the exception of the section on line audits, little evidence was provided concerning the Coding Forms' content validity. Nevertheless, the results obtained by Khatwa and Helmreich provide preliminary evidence of converging validity with the HFACS. Specifically, Khatwa and Helmreich found that the five most frequently identified causal factors accounted for a majority (71%) of the accidents. All of these causal factors concern the flight crew, and none involved supervisory or organizational influences. Similar results were observed by Pape and colleagues (Pape, Wiegmann, & Shappell, 2001).

Uses/Utility. Khatwa and Helmreich identified a number of recommendations for minimizing accidents during the approach and landing phases of flight. Specific recommendations were developed for regulatory agencies, air traffic controllers, air carrier operators, flight crews, manufacturers, and the industry at-large. Unfortunately, we have been unable to locate any follow-up studies that have documented the implementation or effectiveness of these recommendations.

Strengths and Weaknesses. The Coding Form has a number of desirable qualities. First, it identifies both the primary and contributing causes of accidents. While several other systems model the multivariate nature of accidents, the Coding Form explicitly distinguishes between primary and contributing causes. A second strength is that the Coding Form explicitly codes the consequences of each accident or incident. This is extremely valuable, because distinguishing between those combinations of factors that led to catastrophe versus those that did not can provide important insights for developing interventions. A third strength is that the Coding Form requests information regarding how to prevent similar accidents in the future.

Notwithstanding, the Coding Form has its limitations. One drawback is its exceptional length. To ensure that line pilots would properly complete it, the Coding Form would need to be substantially shortened prior to operational use. A second drawback is that it does not provide a formal mechanism for specifying the chain of events. Finally, because the it is devoted to issues involving the approach and landing phases of flight, the Coding Form would need to be expanded to address Human Factors issues in other phases of flight.

University of Texas Line/LOS Checklist Version 4 (LLC4)

Mission Statement. The Line/LOS Checklist Version 4 (LLC4) was perhaps the first widely used set of behavioral markers for assessing Crew Resource Management (CRM) skills.

More recently, it has been used to help carriers define the scope of their CRM training programs. Sections from the LLC4 have also been adapted for use in non-jeopardy Line Operational Safety Audits (LOSAs).

Theoretical Orientation. The performance dimensions included on the LLC4 were originally culled from NTSB accident reports that identified Human Factors issues as either the primary or contributing causes of accidents (Connelly, 1997). Recent versions of the LLC, such as those used in conducting LOSAs, incorporate elements from Reason's (1990) model of human error. This newer version of the LLC has been adapted to assess environmental threats, pilot errors, and the countermeasures that pilots take to mitigate these errors (Helmreich, Klinect, & Wilhelm, 1999).

Data Organization and Management. The LLC4 performance dimensions are organized into six categories: team management and crew communication, situational awareness and decision-making, automation management, special situations, technical proficiency, and overall crew performance. Trained assessors collect ratings at four points during the flight (pre-departure, take-off and climb, cruise, and descent/approach/landing) using a 4-point rating scale with anchors ranging from "poor" (1) to "outstanding" (4). The LLC4 also includes a series of demographic items, including questions that assess the crewmembers' experience flying together. Finally, space is provided for evaluator comments regarding the crew's performance (Helmreich, Butler, Taggart, & Wilhelm, 1995). Many of the descriptive/demographic fields are stored as nominal variables, possibly with pre-defined response options. The behavioral markers are stored as numerical variables, and that the assessor comments are stored as text.

Validation Evidence. Over the years, various iterations of the LLC have been used to assess crew performance using videotaped vignettes and non-jeopardy line audits.

Unfortunately, information concerning the LLC's reliability and validity is not consistently reported from study to study. As a result, it is difficult to make broad generalizations about the instrument's psychometric properties.

In one study (Law & Sherman, 1995), a group of 34 pilot instructors observed videotaped examples of above- and below-average crews performing a simulated flight. The instructors then evaluated the crews' technical proficiency and overall effectiveness using the LLC4. The level of inter-rater agreement ranged between .78 and .94 for overall technical proficiency and between .82 and 1.00 for overall crew effectiveness. While impressive, this high level of agreement may have been artificially inflated by the small variance in the observed ratings. This lack of variance could be due to a number of factors, such as the effectiveness of the rater training, the quality of the instrument, or "true" level of proficiency displayed by the videotaped crews. Unfortunately, these confounds render it impossible to determine the true cause of the results.

Studies that have used the LLC4 in line audits have shown statistically significant mean differences between fleets within the same airline (Helmreich, Wilhelm, Gregorich, & Chidester, 1990), and between airlines (Law & Wilhelm, 1995) on various behavioral markers. However, the observed differences tend to be small, with mean differences averaging between one-third and one-fourth of a scale point (Helmreich & Foushee, 1993).

Strengths and Weaknesses. A major strength of the LLC4 is that it requires the assessor to make separate judgments of the crews' performance for each of the four major segments of flight (i.e., pre-departure, take-off and climb, cruise, descent and landing). This is important, because it helps to delineate the chain of events. Another strength is that the LLC4 collects information regarding the pilots' experience flying together. This is especially important, because teamwork behaviors should evolve as crewmembers become more familiar with one another's typical work styles. As a result, it may be possible to compare and contrast the types of errors made by "recently-paired" crews versus "more established" crews.

Uses/Utility. Perhaps the greatest utility of the LLC4 has been its widespread adoption at a variety of carriers, where it has been used as a common metric for assessing crew performance. A second strength of the LLC4 has been its use in non-jeopardy line audits. Although the process of collecting non-jeopardy line audit data may not provide a completely accurate picture of crew behaviors under typical line conditions, it clearly provides a more accurate picture of crew behavior than that observed under simulated conditions such as Line Oriented Flight Training (LOFT).

Data Collection Tools

In this section, we describe several data collection tools that were developed to help carriers collect, organize, and analyze their ASAP-style incident reports. These systems have built-in taxonomies for classifying Human Factors issues in ASAP reports. As a result, they are well suited to identifying the specific problems faced by an individual carrier and for suggesting targeted interventions. These systems are also important because they can reduce the administrative burdens associated with managing an ASAP program, thereby providing ASAP staff with additional time for data analysis, problem identification, and problem resolution.

University of Texas demonstration ASAP Incident Reporting Form

Mission Statement. The University of Texas' demonstration Incident Reporting Form collects detailed information regarding the demographic, environmental, and behavioral correlates of aviation-related incidents. It was developed in consultation with the Air Transport Association's Human Factors Committee to elicit pilots' perceptions of why an event occurred and what could have been done to prevent it, rather than merely describing what happened during the event (Helmreich & Merritt, 1999).

Theoretical Orientation. The Reporting Form was developed, in part, based on Reason's (1990) model of human error (Helmreich, personal communication, 2002). The Reporting Form explicitly asks the reporter to identify factors that contributed to and mitigated against human error. These factors are grouped into logical categories, such as the cockpit crew, ATC and other constituency groups, aircraft systems and equipment, procedures and documentation, and the external environment.

Data Organization and Management. The Reporting Form includes multiple electronic forms that are linked together by a common set of identifying fields. The first form was designed to be completed by the pilot crew. This form contains seven major sections. These include: event demographics, crewmember demographics, external conditions, the event itself (i.e., a description of the event), contributing causes, recommended improvements, and pilot feedback. Most of these fields are stored as nominal variables with pre-defined response options. A limited number of fields, including the event description and pilot feedback, are stored as text.

The second form is designed to be completed by the carrier's ASAP administrator. This form collects information regarding the event's actual consequences (as determined by company personnel), its severity, and its perceived probability of reoccurrence. This form also permits the ASAP administrator to perform various administrative functions, such as notifying the reporter that the report has been received, and submitting a de-identified report to ASRS.

The third form is designed to be completed by the Webmaster. This form collects additional information that was not included in the first form because of space or time constraints. The Webmaster completes these fields (e.g., hours flying time in the last 90 days) by accessing the reporter's information from company databases (Wilhelm, Klinect, & Jones, 2000).

Validation Evidence. To date, little empirical research has been publicly disseminated concerning the Reporting Form. Prior to January 2002, the form was available online (http://www.psy.utexas.edu/psy/helmreich/ASAPFORM/pilot/pilot.htm). However, it has since been removed. As a result, validation evidence is difficult to locate.

Uses/Utility. We have been unable to identify any studies that have explicitly analyzed and reported data collected using this instrument. However, the Reporting Form has been frequently used as a debriefing tool for analyzing case studies and teaching Crew Resource Management principles (Jones & Tesmer, 1999).

Strengths and Weaknesses. The Reporting Form has a number of strengths. First, it automatically links information from pilot reports, ASAP administrator reports, and Webmaster reports, thereby providing a more complete description of the event in question. Second, the Reporting Form attempts to identify the factors that contributed to and mitigated against the incident. Finally, the Reporting Form collects suggestions from pilots regarding how to prevent the problem's reoccurrence. However, it is not without limitations. The most prominent is its relative lack of emphasis on describing crew processes. Despite collecting a wealth of information on the incident's behavioral and environmental correlates, most of the questions are designed to address "what" happened. Virtually all of the information concerning "why" things happened (e.g., Human Factors issues) are still relegated to the text narrative.

British Airways Safety Information Systems (BASIS)

Mission Statement. The British Airways Safety Information System (BASIS) is a commercial software tool that was designed to help air carriers assess and manage the risks associated with flight operations. BASIS was originally developed to gather and analyze pilot incident reports internally at British Airways. Over time, the BASIS software has been expanded and marketed to other organizations. Currently, BASIS can integrate data from multiple information sources – including incident reports (e.g., submitted by pilot crews, cabin crews, maintenance crews, mechanics), FOQA exceedences, and line audits – thereby providing a holistic picture of an air carrier's safety health (http://www.winbasis.com). BASIS can also produce statistical summaries and graphs of risk factors, which can be sorted by their frequency of occurrence or by subject matter experts' assessments of their severity. A separate BASIS module allows participating carriers to share de-identified data with other BASIS customers via a secure Internet database.

Theoretical Orientation. Because BASIS is proprietary, information regarding its development remains confidential. As a result, it is difficult to determine what theoretical models, if any, were used to guide its development. Because there is no mention of any particular theoretical model in the BASIS product literature or instruction manuals, we speculate that the BASIS taxonomy was not developed based on any particular theory of human performance or human error.

Data Organization and Management. As noted earlier, because BASIS is proprietary, information regarding its how the various types of data are organized, linked, managed, and summarized is not available. However, BASIS is capable of handling both quantitative and qualitative data, and the BASIS taxonomy can differentiate between causes and effects (http://www.winbasis.com/products/basis_prod_asr.htm).

Validation Evidence. To date, we have been unable to locate any studies that assess the reliability or validity of BASIS Air Safety Reporting taxonomy.

Uses/Utility. BASIS is currently being used by over 100 organizations, including air carriers, pilot unions, trade associations, and aircraft manufacturers to manage the risks associated with flight operations. The BASIS Safety Information Exchange feature currently maintains a database of over 43,000 incidents that were compiled from BASIS-participating carriers. However, we have been unable to locate any studies, published or unpublished, that have analyzed BASIS incident reports for scientific purposes.

Strengths and Weaknesses. Based on our review, the BASIS program has a number of strengths. The first is that it allows carrier personnel to search for trends across ASAP-style incident reports, line audit data, and FOQA exceedences. A second strength is its use of a structured "interview" format that elicits standardized responses from line pilots regarding the "Who?," "What?," "Where?," "When?," "Why?," and "How?" details of the incident. This information is then drawn into a Human Factors Report module that allows the user to diagram the chain of events (http://www.winbasis.com/products/basis_prod_hrf.htm). A third strength is that BASIS allows users to share de-identified data with other organizations, so that carriers can learn from each other's mistakes (http:// www.winbasis.com/products/basis_prod_sie.htm). A final strength is the capacity for subject matter experts to assign a risk judgment to each incident report based on its perceived severity and probability of reoccurrence.

Conclusions

Although each system took a slightly different approach to collecting, analyzing, and interpreting safety-related data, all have made meaningful contributions to improving air safety. The subtle differences among them derive largely from their unique mission statements, theoretical orientations, target populations, and logistical constraints. Perhaps more than anything else, our review has taught us that no system can be evaluated independently of these considerations.

In this section, we draw upon the best practices uncovered during our review to develop a wish list of the fields that we think an "ideal" ASAP taxonomy should include, and the features that an "ideal" electronic reporting form and analysis tool should perform. As mentioned earlier, our goal is to develop a generalized Human Factors taxonomy and electronic data reporting/analysis tool that will be provided at no cost to industry. We hope that pilots will use this system to report their errors, flight safety staff will use this system to identify safety-related trends, training personnel will use this system to develop new training programs, and researchers will use this system to better understand the nature of pilot error (Beaubien & Baker, under review). We will also encourage carriers to modify this tool, as necessary, to best suit their unique culture and mission-specific requirements.

This system is "ideal" in the sense that when fully-functional, it will address the ASAP-specific needs that we have identified through extensive interactions with line pilots, ASAP administrators, and government regulators. Given our unique considerations, we believe that our ideal system must:

1. Be comprehensive in its treatment of Human Factors issues;
2. Be user-friendly, even for those with minimal formal training in Human Factors;
3. Reliably classify similar events despite differences in the reporter's verbal skills;
4. Apply equally well to describing the problems faced by pilots at regional and major air carriers;
5. Allow carriers to identify specific causes of incidents and provide specific guidance for resolving them;
6. Facilitate communication among pilots, management, researchers, and regulators;
7. Not place excessive demands on the user;
8. Be generic enough to be adapted for use with other forms of safety-related data.

It is "ideal" only in relation to our project goals. Because it is still a work-in-progress, we have no evidence to support our assertion that any ASAP system – even with these characteristics – will allow ASAP data to be used to its fullest potential of identifying safety trends, developing data-driven interventions, and empirically assessing the effectiveness of these interventions. However, once the system becomes operational, we intend to conduct follow-up studies to assess the extent to which this system achieves our goals.

Developing an "Ideal" Taxonomy for Classifying Human Factors Data in ASAP

In this section, we describe our current plans for classifying Human Factors issues in ASAP. One of our project's long-term goals is to eventually integrate ASAP data with other forms of safety-related data, such as Advanced Qualification Program (AQP) performance ratings and line audit summaries. In a perfect world, this would allow carriers to search for trends by means of triangulation. To this end, our initial plan for coding ASAP data is to keep the taxonomy as simple as possible. Later, we can create more complex data fields by using "recode" commands and "concatenate" functions. An example appears below.

We envision that the Human Factors data would be organized into one or more lists of exemplars, each of which should be only one level "deep." For example, crew processes would be described using a generic list of teamwork behaviors (e.g., "communication," "coordination," "decision-making," and so forth). Each list could accept multiple responses. For example, if the pilot reported an error chain that contained 3 "links," then 3 separate teamwork behaviors could be selected.

For each Human Factors issue that is selected, supplemental information would be collected in separate variables known as "reason codes." The reason codes would be used to identify relevant information, such as the order of occurrence (e.g., in the chain of events), relevance (i.e., primary vs. contributory cause), relationship with others (e.g., co-pilot, dispatch, ATC, maintenance), and other relevant factors. The information would be stored in this manner because it is extremely flexible for use in conducting statistical analyses.

For example, a typical ASAP report might indicate that the crew misunderstood an ATC directive during the final approach. As a result, they lost situation awareness and made an incorrect decision to land on the wrong runway. Using the method described above, the respondent would identify "communication" as a crew process error. The responding pilot would then indicate the error's relationship to others (i.e., it involved ATC), its relevance (i.e., it was a contributory cause), and its order of occurrence (i.e., it was first error in a three-error chain). Finally, the responding pilot would describe the second and third errors – situation awareness and decision-making, respectively – using the same procedure.

Because each reason code is stored as a separate variable, the data can be combined in a variety of ways. For example, "communication" and "involving ATC" can be combined to form "communication involving ATC". This will permit carriers to easily compare their ASAP data with other forms of safety-related data, such as AQP performance ratings and LOSA audit summaries.

Developing an "Ideal" ASAP Reporting Form

In this section, we describe our current plans for developing an ASAP reporting form. Although this is something of an oversimplification, there are two overarching issues to consider when developing an ASAP reporting form. The first issue involves deciding what types of information to collect. Based on our review, we have identified five broad categories of information that should be included in an ideal ASAP reporting form. These categories include: crewmember and flight demographic information (e.g., seat position, flying experience, flight number, origin and destination, etc.), antecedent conditions (e.g., weather, meteorological conditions, air traffic, etc.), Human Factors information (e.g., crew processes, automation issues, human physiological limitations, etc.), consequences/outcomes (e.g., loss of control, runway incursion, fire, etc.), and lessons learned (e.g., suggestions for preventing similar occurrences, an assessment of the incident's safety implications, etc.). All five categories are essential for understanding not only how the problem was caused, but how to prevent it from re-occurring.

The second major issue involves organizing this information as effectively as possible. Drawing on the BASIS system, an ideal reporting form might be organized much like a structured interview which guides the respondent using a logical progression of questions (e.g., "Who?," "What?," "Where?," "When?," "Why?," and "How?"). Carefully designing the user interface to minimize workload, and refining it based upon the results of usability tests, should lessen the amount of training required for pilots to use the system effectively. We also envision that the reporting form would be linked to other carrier databases. This would offer a number of advantages. For example, the reporting pilot could simply enter his/her personal identification number into the form, and all relevant demographic fields (e.g., name, position, flying experience) would automatically display. Automation of this type has the potential to reduce typographical and memory-based errors. It can also decrease the amount of time required to complete the demographic fields, thereby freeing up additional time for the reporting pilot to describe more substantive Human Factors issues.

Developing an "Ideal" ASAP Data Collection and Reporting Tool

In this section, we describe our current plans for developing a computerized ASAP data collection, management, and analysis tool. Based on our review, the University of Texas (UT) Demonstration Incident Reporting Form provides a good example of the functions that an ideal

ASAP reporting database should perform. We envision that an ideal ASAP database would include separate forms for data entry, system maintenance, and data analysis. Each form would be tailored to the needs of a separate audience. The reporting pilots would complete the first form. This form would collect information such as event demographics, external conditions, the event itself (i.e., a description of the event), contributing causes, and recommendations for preventing the event's reoccurrence.

The second form would be completed by the carrier's ASAP administrator or Event Review Committee (ERC). As before, this form would collect information regarding the actual consequences (as determined by company personnel) of the event, an official assessment of the incident's severity and probability of reoccurrence, and an official estimate of the perceived difficulty required to remedy the problem. However, the ASAP reporting tool would also automate a number of administrative functions, such as notifying the reporting pilot that the report has been received, de-identifying the report, and submitting the report to ASRS Finally, we envision that the reporting tool would also calculate simple charts and graphs (e.g., frequency distributions and cross-tabulation tables) of the most frequently occurring types of problems, and export the data for more sophisticated statistical analyses.

Recommendations for Practice

Our review highlighted a number of other issues that need to be considered in error collection, reporting, and analysis. In practice, we believe that the taxonomy and reporting form must be made available to pilots prior to them experiencing an incident. We believe that if pilots have the taxonomy and incident reporting forms available prior to experiencing an incident, they can better organize their thinking when submitting their report. In the best case scenario, training in how to use the system would be integrated into carriers' existing Human Factors or Crew Resource Management (CRM) training courses.

We also believe that all ASAP reporting systems should collect incident reports from multiple constituencies, such as pilot crews, cabin crews, and maintenance crews. Moreover, we believe that each constituent group should receive frequent (e.g., monthly) feedback in the form of a newsletter. This newsletter should contain examples of high priority issues that occurred during the previous month, example (de-identified) ASAP reports, suggestions for preventing a re-occurrence in the future, and links/citations to relevant carrier and FAA aviation regulations.

Finally, we believe that ASAP programs need to be thoroughly documented, with all relevant documentation contained in a centralized repository. During our review, we encountered several systems that initially seemed promising, but were largely undocumented. As a result, it was difficult to determine how their data were collected, organized, or analyzed. Therefore, we are developing a web site that will eventually provide users with up-to-date information regarding the taxonomy and its associated research (e.g., technical reports, research protocols, instruments, briefing slides, etc.).

References

Baker, D. P., Chidester, T. R., & Brannick, M. T. (1998). *Developing scenario event sets for LOE: A data-driven approach*. Unpublished report. Tampa, FL: The University of South Florida.

Beaubien, J. M., & Baker, D.P. (under review). *Toward a generalized Human Factors taxonomy for classifying ASAP incident reports, AQP performance ratings, and FOQA output*.

CHIRP Charitable Trust (1999). CHIRP Survey – 1999, *Feedback, 51*. Available: http://www.chirp.co.uk/air_transport/FB51_Supplement.htm.

Connelly, E. P. (1997). *A resource package for CRM developers*. Technical Report 97-03. Houston, TX: University of Texas Aerospace Crew Research Project.

Creswell, J. W. (1994). *Research design*. Thousand Oaks, CA: Sage.

Federal Aviation Administration. (2000). *Aviation Safety Action Programs (ASAP)*. Advisory Circular 120-66A. Washington, DC: Author

Federal Aviation Administration. (1987). *Aviation Safety Reporting Program*. Advisory Circular 00-46D. Washington, DC: Author.

Fleishman, E. A., & Mumford, M. D. (1991). Evaluating classifications of job behavior: A construct validation of the ability requirement scales. *Personnel Psychology, 44*, 523-575.

Fleishman, E. A., & Quaintance, M. K. (1984). *Taxonomies of human performance: The description of human tasks*. Orlando, FL: Academic Press.

Fleiss, J. (1981). *Statistical Methods for Rates and Proportions*. New York: Wiley.

Global Aviation Information Network. (2001, December). *Guide to methods and tools for airline flight safety analysis*. Available: http://www.gainweb.org/.

Helmreich, R. L., Butler, R. E., Taggart, W. R., & Wilhelm, J. A. (1995). *The NASA/University of Texas/FAA Line/LOS checklist: A behavioral marker-based checklist for CRM skills assessment*. Technical paper 94-02. Houston, TX: University of Texas Aerospace Crew Research Project.

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

Helmreich, R. L., Klinect, J. R., & Wilhelm, J. A. (1999). Models of threat, error, and CRM in flight operations. In *Proceedings of the 10th International Symposium on Aviation Psychology*, (pp 677-682). Columbus, OH: The Ohio State University.

Helmreich, R .L., & Merritt, A. C. (1999). *Culture at work in aviation and medicine: National, organizational and professional influences*. Aldershot, UK: Ashgate.

Helmreich, R. L., Wilhelm, J. A., Gregorich, S., & Chidester, T. (1990). Preliminary results from the evaluation of cockpit resource management training: Performance ratings of flight crews. *Aviation, Space, and Environmental Medicine, 61*, 576-579.

Ingebretson, E. A., & Morrison, R. (1999). *Aviation Data Sources with Human Factors Information*. Unpublished technical memorandum. Moffett Field, CA: NASA-Ames Research Center.

Jentsch, F., Barnett, J., Bowers, C. A., & Salas, E. (1999). Who is flying this plane anyway? What mishaps tell us about crew member role assignment and air crew situation awareness. *Human Factors, 41*, 1-14.

Jones, S. G., & Tesmer, B. (1999). A new tool for investigating and tracking human factors issues in incidents. In *Proceedings of the 10th International Symposium on Aviation Psychology* (pp 696-701). Columbus, OH: The Ohio State University.

Khatwa, R., & Helmreich, R. L. (1998). Analysis of critical factors during approach and landing in accidents and normal flight. *Flight Safety Digest (November 1998 - February 1999)*, 1-92.

Khatwa, R., & Roelen, A. L. (1997). Controlled flight into terrain (CFIT) accidents of air taxi, regional and major operators. In *Proceedings of the 9th Flight Safety Foundation European Safety Seminar*. Alexandria, VA: Flight Safety Foundation.

Law, J. R., & Sherman, P. J. (1995). Do raters agree? Assessing inter-rater agreement in the evaluation of air Crew Resource Management skills. In *Proceedings of the 8th International Symposium on Aviation Psychology* (pp. 608-612). Columbus, OH: The Ohio State University.

Law, J. R., & Wilhelm, J. A. (1995). Ratings of CRM skill markers in domestic and international operations. In *Proceedings of the 8th International Symposium on Aviation Psychology* (pp. 669-675). Columbus, OH: The Ohio State University.

Lee, R. (2001). *New directions in air safety*. Unpublished report. Civic Square, Australian Capital Territory (ACT): Bureau of Air Safety Investigation.

National Aeronautics and Space Administration. (1999). *Aviation Safety Reporting System program overview*. Available: http://asrs.arc.nasa.gov/briefing_nf.htm#2.

Pape, A. M., Wiegmann, D. A., & Shappell, S. A. (2001). Air traffic control (ATC) related accidents and incidents: A human factors analysis. In *Proceedings of the 11th*

International Symposium on Aviation Psychology. Columbus, OH: The Ohio State University.

Reason, J. (1990). *Human error*. New York: Cambridge University Press.

Sarter, N., & Alexander, H. (2000). Error types and related error detection mechanisms in the aviation domain: An analysis of aviation safety reporting system incident reports. *International Journal of Aviation Psychology, 10*, 189-206.

Shappell, S. A., & Wiegmann, D. A. (2000). *The Human Factors Analysis and Classification System – HFACS*. Technical Report DOT/FAA/AM-00/7. Washington, DC: Office of Aviation Medicine.

Stone, R. B., Babcock, G. L., & Edmunds, W. W. (1985). Pilot judgment: An operational viewpoint. *Aviation Space and Environmental Medicine, 56*, 149-152.

U. S. Navy. (2001, March). *OPNAV Instruction 3750.6R, Appendix J*. Washington, DC: Office of the Chief of Naval Operations.

Wiegmann, D. A., & Shappell, S. A. (2001). *A human error analysis of commercial aviation accidents using the Human Factors Analysis and Classification System (HFACS)*. Technical Report DOT/FAA/AM-01/3. Washington, DC: Office of Aviation Medicine.

Wilhelm, J. A., Klinect, J. & Jones, S. (2000). *The University of Texas Human Factors Research Project Demonstration ASAP Incident Reporting Form*. Available: http://www.psy.utexas.ed/psy/helmreich/ASAPFORM/ASAPEXPLN.htm.

Table 1. Comparison and Contrast of Human Factors Taxonomies, Reporting Systems, and Data collection Tools (Taxonomic Issues)

	ASRS	CHIRP	CAIR	HFACS	Approach and Landing Coding Form	LLC4	Univ. of Texas ASAP Form	BASIS
Developed based on a theoretical model of human performance or human error	No	No	No	Yes	Yes	Yes	Yes	No
Provides guidance for reporting or interpreting the chain of events	Yes	No	No	No	No	No	No	Yes
Elicits suggestions for preventing a re-occurrence	Yes	No	Yes	No	Yes	Yes	Yes	Yes
Distinguishes between the incidents' primary and contributing causes	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes
Report is coded by multiple analysts	Yes	Unknown	Unknown	Yes	Yes	No	No	No
Distinguishes between errors of omission and commission	Yes	No	No	Yes	No	No	No	Unknown
Explicitly considers supervisory and organizational factors	No	No	No	Yes	Yes	Yes	Yes	Unknown
Organized in a hierarchical structure	No	No	No	Yes	Yes	No	No	Unknown
Specific enough to identify specific problems and suggest specific interventions	No	No	No	No	Yes	No	No	Unknown
Identifies the incident's consequences or outcomes	No	No	No	No	Yes	No	No	Yes

Note: This table was compiled using publicly-available information. Information from unpublished or limited-circulation documents may not be reflected herein.

Table 2. Comparison and Contrast of Human Factors Taxonomies, Reporting Systems, and Data collection Tools (Database Issues)

	ASRS	CHIRP	CAIR	HFACS	Approach and Landing Coding Form	LLC4	Univ. of Texas ASAP Form	BASIS
Provides separate data entry forms for pilots, ASAP administrators, and webmasters	Yes	Yes	No	N/A	N/A	N/A	Yes	Unknown
Automatically de-identifies the incident report	Yes	Yes	Yes	N/A	N/A	N/A	No	Yes
Automatically files an ASRS report	N/A	N/A	N/A	N/A	N/A	N/A	No	N/A
Capable of producing statistical graphs and charts	No	No	No	N/A	N/A	N/A	No	Yes
Allows Subject Matter Experts (SMEs) to include judgments (e.g., threats to safety, difficulty to remedy)	No	No	No	N/A	N/A	N/A	No	Yes

Note: This table was compiled using publicly-available information. Information from unpublished or limited-circulation documents may not be reflected herein.

Table 3. Comparison and Contrast of Human Factors Taxonomies, Reporting Systems, and Data collection Tools (Miscellaneous Issues)

	ASRS	CHIRP	CAIR	HFACS	Approach and Landing Coding Form	LLC4	Univ. of Texas ASAP Form	BASIS
Provides tailored feedback to individual constituency groups	Yes	Yes	Yes	N/A	N/A	N/A	No	No
Codebook, taxonomy, or glossary is made available to the public	No	Yes	No	Yes	Yes	Yes	Yes	No
Data collection form is organized in a structured interview guide	No	No	No	No	No	No	No	Yes

Note: This table was compiled using publicly-available information. Information from unpublished or limited-circulation documents may not be reflected herein.

Author Notes

This research was supported through grant 99-G-048 from the FAA's Office of the Chief Scientific and Technical Advisor for Human Factors (AAR-100), Dr. David P. Baker, Principal Investigator.

Correspondence concerning this article should be addressed to J. Matthew Beaubien, American Institutes for Research, 1000 Thomas Jefferson Street NW, Washington DC, 20007. E-mail: jbeaubien@air.org.