Beaubien, J.M., & Baker, D.P. (2003). Toward a generalized human factors taxonomy for classifying ASAP incident reports, AQP performance ratings, and FOQA output. In R.S. Jensen (Ed.), *Proceedings of the 12th International Symposioum on Aviation Psychology*. Dayton, OH: Wright State University Press.

TOWARD A GENERALIZED HUMAN FACTORS TAXONOMY FOR CLASSIFYING ASAP INCIDENT REPORTS, AQP PERFORMANCE RATINGS, AND FOQA OUTPUT

J. Matthew Beaubien
David P. Baker
American Institutes for Research
Washington, DC

Over the years, the FAA has partnered with industry to develop a number of programs for reporting, classifying, and analyzing safety-related data. Despite their successes, none of these programs has been able to integrate data from multiple sources. To address this problem, we are developing a generalized Human Factors taxonomy for classifying de-identified ASAP incident reports, AQP performance ratings, and FOQA output. Eventually, this taxonomy will be embedded into a series of searchable computer databases that "speak a common language," thereby allowing the search for trends.

Introduction

During the past 20 years, the FAA has partnered with industry to collect safety-related information from a variety of sources, including incident reports, simulated training flights, and flight data recorders. Collectively, these programs have done a great deal to increase the margin of safety. In the sections below, we describe these programs in greater detail.

Aviation Safety Action Program (ASAP)

The Aviation Safety Action Program is a voluntary, non-punitive incident reporting system. Like NASA's Aviation Safety Reporting System (ASRS), ASAP encourages pilots to self-report their errors by providing limited immunity from prosecution. However, unlike the ASRS, ASAP is carrier-specific (Federal Aviation Administration, 2000a).

Drawing on the ASRS example, many carriers' ASAPs collect information that describe the reporting pilot (e.g., seat position, flying time), relevant flight conditions at the time of the event (e.g., weather, phase of flight), and the event's outcome (e.g. runway incursion, loss of separation). Space is also provided so that the reporting pilot can write a short narrative describing the causal factors that precipitated the event, as well as suggestions for preventing its re-occurrence.

De-identified ASAP reports are typically stored in a relational database as a series of alphanumeric fields. As a general rule, most of the Human Factors issues can be found within the text narratives. Because these narratives require costly and time-consuming content analysis, ASAP data has rarely been used to its fullest capacity when developing training objectives, LOS scenarios, and other safety-related interventions.

Advanced Qualification Program (AQP)

The Advanced Qualification Program is a voluntary alternative to traditional pilot training under 14 CFR Part 121. Following an instructional systems design (ISD) approach, AQP training developers begin with a comprehensive task analysis of the technical and Crew Resource Management (CRM) requirements for each fleet. The results of these task analyses are then used to guide the development of fleet-specific training programs and pilot evaluation worksheets (Federal Aviation Administration, 1991).

AQP performance ratings are typically collected during Line Operational Flight Training (LOFT) that mimics normal and emergency flight operations. De-identified performance data are typically stored in a relational database as a series of alphanumeric fields. Several fields are used to describe the crews' training experiences (e.g., LOFT identification number, LOFT instructor's name, date of LOFT). The database also contains a list of behaviors that were evaluated during the LOFT, as well as a rating of the crew's proficiency on each behavior. Finally, the database links each behavior to the training objective that it is intended to measure.

Flight Operations Quality Assurance (FOQA)

Flight Operations Quality Assurance is a voluntary initiative for collecting and analyzing flight data recorder (FDR) output from routine flights. Because FOQA data does not rely on the errors or biases of a human reporter, it is completely objective.

FOQA data are downloaded from the FDR as a digital data stream of over one hundred flight parameters (e.g., speed, direction, flap settings) that are continually measured throughout the flight. Specialized software then translates the data stream into "exceedences" or deviations from acceptable threshold values (e.g., a descent rate that exceeds 1000 feet per minute on approach).

All exceedences, which are derived from the carrier's operating standards and the manufacturer's operating limitations, are weighted in terms of their severity (Federal Aviation Administration 2000b). FOQA data are typically stored in a relational database as a series of numeric fields. After being verified for accuracy, FOQA data are de-identified.

The Integrated Flight Quality Assurance System (IFQASys)

Each of these programs has done a great deal to improve the margin of safety. Despite their successes, however, none have been able to integrate data from multiple sources. To address this problem, we are developing a generalized Human Factors taxonomy for classifying de-identified ASAP incident reports, AQP performance ratings, and FOQA output. Eventually, this taxonomy will be embedded into a series of searchable computer databases that "speak a common language."

Although individual records will be de-identified, carriers will be able to identify safety-related problems by triangulation. For example, if a carrier's ASAP reports indicate that non-precision approaches are a problem, their FOQA and AQP data can be analyzed to verify the problem's existence.

The value-added benefits of this project include: the capacity to identify problems by triangulation, the capacity to rank order safety-related problems by frequency of occurrence or perceived probability of re-occurrence, the capacity to develop data-driven interventions, and the capacity to measure the effectiveness of these interventions by statistically comparing pre- and post-intervention data.

This project is being developed under the IFQASys model (Longridge, 2002). IFQASys was founded on the belief that the FAA can no longer achieve further safety improvements through enforcement action. Rather, IFQASys calls for the voluntary sharing of de-identified data between airlines, between airlines and industry, and between industry and government.

According to the IFQASys model, industry defines how the system will evolve. For example, the type and amount of data to be shared will ultimately be determined by the airlines, not by the FAA. The FAA's primary role in this project is to provide technical guidance for industry initiatives and to fund the development of tools for analyzing trend data (Longridge, 2002).

Project Goals

Our project has three primary goals. The first goal is to develop a comprehensive ASAP taxonomy that will allow carriers to quantify the Human Factors issues that their crews face during typical line operations. Armed with this information, carriers can then develop data-driven interventions and evaluate the effectiveness of these interventions by statistically comparing pre- and post-intervention data.

Our second goal is to embed this taxonomy within a searchable data collection and reporting tool. Doing so will streamline the process of collecting, managing, and reporting ASAP data. As a result, the carriers' limited resources can be devoted to more goal-directed tasks such as problem identification, analysis, and resolution.

Our third goal is to extend all or part of this proposed taxonomy to include de-identified AQP performance ratings and de-identified FOQA output. The end result will be three separate databases that use a common Human Factors taxonomy. Although individual records will be de-identified, carriers will be able to identify safety-related problems by triangulation. For example, if a carrier's ASAP reports indicate that non-precision approaches are a problem, their FOQA and AQP data can be analyzed to verify the problem's existence.

The value-added benefits of this project include: the capacity to identify problems by triangulation, the capacity to rank order safety-related problems by frequency of occurrence or perceived probability of re-occurrence, the capacity to develop data-driven interventions, and the capacity to measure the effectiveness of these interventions by statistically comparing pre- and post-intervention data.

The Reasons for Starting with ASAP

After reviewing all three programs, we chose to begin our taxonomy development work using ASAP data. This decision was made for several reasons. Practically speaking, ASAP contains the most detailed information for determining "why" an event occurred. By focusing on ASAP as our starting point, we were able to maintain our focus on answering the question "why?" from the outset.

Other reasons were based purely on logistical constraints. Simply put, we knew that the text-based nature of ASAP would lend itself easily to developing a Human Factors taxonomy. We also know that a sizeable number of airlines already have ASAP or ASAP-like programs. Therefore, it is our intention to field test the ASAP taxonomy at a number of carriers as we will begin work on translating the taxonomy for use in AQP and FOQA. As the field tests proceed, we hope to apply the lessons learned from ASAP to our new work.

Like all research projects, we started by conducting a thorough review of existing Human Factors taxonomies, accident/incident reporting systems, and data collection tools (Beaubien & Baker, in press). After completing our literature review, we realized that no one system embodied all the features that we needed. Therefore, we decided to design our own tool from scratch. We also realized that our taxonomy – and the resulting database – must meet certain requirements if it is to be used on a daily basis for identifying specific operational problems and suggesting interventions to address those problems. In this section, we outline several of these constraints.

First, the taxonomy must be comprehensive in its treatment of Human Factors issues. Simply put, the taxonomy must address more than just Crew Resource Management (CRM) issues. It must also address human error (e.g., memory slips, lapses of attention, error countermeasures, etc.), human-automation interaction (e.g., mode awareness errors, automation-induced complacency, etc.), physiological limitations (e.g., fatigue, rest-duty cycles, jetlag, etc.), and the like. Failing to include such topics would result in a product that has limited real-world application.

Second, the taxonomy must be user-friendly, even for those who have minimal formal training in Human Factors. For example, we intend to embed our taxonomy within an electronic ASAP reporting form, so that pilots can file their reports on-line. Therefore, it is imperative that the taxonomic structure be consistent with how pilots mentally organize these Human Factors issues.

Third, the taxonomy must reliably classify similar events despite minor differences in the wording of the narrative text. Quite simply, if the taxonomy is not robust enough to reliably classify similar events, then it will be impossible to draw valid conclusions when aggregating the data for statistical analysis.

Fourth, the taxonomy must apply equally well to describing the problems faced by pilots at regional and major air carriers. Failing to include both groups in our research will alienate a sizeable segment of the pilot population. Moreover, developing a tool for only a single constituency group will make it impossible to compare and contrast these groups. Such information is essential, because interventions developed for regional carriers may not necessarily generalize to major carriers, and vice-versa.

Fifth, the taxonomy must help carriers to identify specific problems and provide specific recommendations for resolving them. Quite simply, if the tool does not help carriers to solve problems, then they will not use it. The difficulty here involves deciding how much detail to include in the taxonomy. Specifically, we will need to maintain a balance between identifying (and solving) carrier-specific problems and systemic problems that occur throughout the national aerospace system.

Sixth, the taxonomy must facilitate communication by using standardized terminology that can be understood by both researchers and practitioners. Researchers and practitioners bring different skill sets to the analysis and resolution of Human Factors problems. Moreover, they must work together to develop effective solutions.

Seventh, the taxonomy must not place excessive demands on the user (e.g., cognitive limitations, time requirements), regardless of whether the user is a pilot who is reporting an event or a researcher who is analyzing a de-identified incident report. If the taxonomy is too burdensome, users will avoid using it. Therefore, one of our goals is to automate much of the work, such as running reports and queries, so that carrier personnel can spend more time identifying, analyzing, and resolving problems.

Finally, the taxonomy must be generic enough that it can be adapted to de-identified AQP performance ratings and FOQA output.

Integrating ASAP, AQP, and FOQA Data

We recognize that integrating the data will not be an easy task. In this section, we describe some of our proposed strategies for developing a generalized Human Factors taxonomy for use in ASAP, AQP, and FOQA.

Integrating ASAP and AQP Data. In a perfect world, fleet managers and AQP training developers would use ASAP incident reports to identify common problems that their pilots experience on the line (e.g., runway incursions). AQP training programs would then be developed to address these problems. After the training had been completed, the ASAP data would be queried to determine if there was a statistically or practically significant decline in the number of ASAP reports on that particular topic.

Our initial plan for integrating ASAP and AQP 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 (e.g., "communication" and "involving ATC" can be combined to form "communication involving ATC"). This will permit carriers to compare their ASAP data with the AQP behaviors that are measured during LOFT and LOE.

Integrating ASAP and FOQA Data. Integrating ASAP and FOQA data will be a considerably more difficult task. Unlike AQP, which collects information on crew behaviors (e.g., crew communications with ATC), FOQA only collects information on outcomes (e.g., airspeed violations, altitude busts). As a result, ASAP and FOQA data can only be compared using outcomes data.

Fortunately, virtually every ASAP reporting form collects information on the event's outcome. In many cases, the outcome is described within the text narrative. However, we recommend that the ASAP report form also contain a list of common outcomes. For ease of use, this list may be sorted by phase of flight.

After describing the event in the narrative text, the reporting pilot would be asked to choose which outcome described the event that they experienced. Summary statistics, such as frequency distributions, could then be computed to identify the most carrier's biggest problems. Alternatively, cross-tabulation tables could be computed to identify common covariates, such as phase of flight or destination airport.

Admittedly, our knowledge of FOQA is limited. However, we have been researching a number of FOQA initiatives such as the Aviation Performance Measuring System (APMS; National Aeronautics and Space Administration, 2001). APMS staff are developing a series of methodologies, software algorithms, and procedures for converting FOQA data streams into fields and variables that are amenable to statistical analysis. We intend to draw upon their work when integrating ASAP and FOQA data.

Developing an "Ideal" ASAP Reporting form

As mentioned earlier, we have chosen to begin our research by focusing on ASAP. Once the ASAP taxonomy has been developed, we will extend it to AQP and FOQA data. In this section, we describe our current plans for developing an "ideal" ASAP data reporting form. 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.

Because this is very much 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.

Although this is something of an oversimplification, there are two overarching issues to consider when developing a 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, air traffic, meteorological conditions, etc.), Human Factors information (e.g., CRM, automation issues, human physiological limitations, etc.), 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 how the problem was caused and how to prevent it from re-occurring.

The second major issue involves organizing this information as efficiently as possible, so that all of the constraints are addressed. Drawing on the BASIS system (www.winbasis.com), we envision that an ideal reporting form might be organized like a structured interview which guides the respondent using a logical progression of questions (e.g., "Who?," "What?," "Where?," "When?," and "Why?").

We also envision that an ideal 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 an "ideal" ASAP data collection and reporting tool. Based on our review (Beaubien & Baker, in press), the University of Texas (UT) Incident Reporting Form (Wilhelm, Klinect, & Jones, 2000) provides a good example of the functions that an ideal ASAP reporting database should perform. Drawing on the UT prototype, 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, crewmember demographics, external conditions, the event itself (i.e., a description of the event), contributing causes, and recommendations for preventing the event's re-occurrence.

The second form would be completed by the carrier's ASAP administrator or Event Review Committee (ERC). 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 re-occurrence, and an official estimate of the perceived difficulty required to remedy the problem. However, an ideal ASAP reporting tool would 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 Reporting tools, such as calculating charts and graphs (e.g., frequency distributions and cross-tabulation tables) of the most frequently occurring problems would also available on this form.

Recommendations for Practice

In addition to the taxonomic and database issues, our review highlighted a number of other issues that do not fit neatly into either category. 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.

An ideal ASAP reporting system should also collect incident reports from multiple constituencies, such as pilot crews, cabin crews, maintenance crews, and ramp personnel. Moreover, 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.

Conclusions

Although there are a number of programs for reporting, classifying, and analyzing safety-related data, none have been able to integrate data from multiple sources. In this concept paper, we have described a multi-year effort to integrate safety-related data from ASAP incident reports, AQP performance ratings, and FOQA output.

By building on small but frequent successes, we hope to encourage industry to not only participate in this project, but to take the lead in visioning how it can be used to solve systemic problems. We invite all interested airlines to contact us to learn more about the project, or to participate in this truly worthwhile endeavor.

Acknowledgments

Funding for this research was provided 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. The views expressed in this paper are those of the authors and should not be construed as an official FAA position, unless so designated by other official documentation.

We would like to thank Capt. Ron Bean, Capt. Ron Whipple, and Ms. Vickie Foster at American Eagle Airlines for helping develop this taxonomy. We would also like to thank Mr. David Neu and Ms. Regina Harris from the Universal Technical Resource Service for developing the computer database and user interface. When complete, the taxonomy and associated databases will be made freely available to industry. Therefore, we welcome additional airline partners in refining and assessing the tool. Interested parties should contact:

J. Matthew Beaubien, Ph.D.
Research Scientist
American Institutes for Research
1000 Thomas Jefferson Street NW
Washington, DC 20007
202-342-5133
jbeaubien@air.org

References

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

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

Federal Aviation Administration. (2000b). *Flight Operational Quality Assurance (FOQA) Program Approval Procedures and Continued Program Monitoring*. Handbook Bulleting for Air Transportation (HBAT) 00-11. Washington, DC: Author.

Federal Aviation Administration. (1991). *Advanced qualification program*. Advisory Circular 120-54. Washington, DC: Author.

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

Longridge, T. L. (2002, March). *Integrated Flight Quality Assurance System (IFQASys): A Flexible Architecture*. Presentation to the ARC Data Sharing Subcommittee.

National Aeronautics and Space Administration. (2001). *Aviation Performance Measuring System*. Available:http://human-factors.arc.nasa.gov/projects/ihs/aviationperf.html.

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.