DATA WAREHOUSING FOR AQP DATABASES

Dr. Robert W. Holt
George Mason University
Fairfax, Virginia

ABSTRACT

The FAA Advanced Qualification Program will create high quality data on pilot performance, instructor and evaluator standardization, and training results which will typically be stored in different databases. For maximum utility for answering the training and operational questions of a fleet, these databases must be effectively interconnected and easily queried. One solution is a Data Warehouse (DW) of interconnected databases. The DW will give information to describe performance, test training effectiveness and answer operational questions.

INTRODUCTION

The Advanced Qualification Program (AQP) of the FAA requires that air carriers collect reliable and valid data. This requirement will result in a large volume of high-quality data organized into databases. Presently, AQP requires two databases, a Program Audit Database (PADB) and a Pilot Performance Database (PPDB). The PADB contains the task analysis, required skills, proficiency objectives, qualification standards, and the curriculum structure, topics, and elements. The PPDB should contain initial qualification results (systems, procedures, and maneuvers validation, Line Operational Evaluation (LOE), Initial Operating Experience line evaluation (IOE)), as well as continuing qualification results (first look and other maneuvers, LOE, line evaluations).

In addition to the required AQP databases, possible carrier databases that would give relevant information include an Instructor/Evaluator Database (IEDB), a company version of the ASRS type of incident reports (CASRS), and a database focused on pilot Knowledge, Skills, and Abilities (PKSA). The IEDB should include information relevant to the quality of instruction and evaluation for each instructor or evaluator. For instructors, instructional qualifications, instructional experience (e.g. classes taught), class evaluations of the instructor, and formal evaluations of the performance of classes taught by the instructor could be included. For evaluators, evaluation experience, the results of evaluation standardization sessions, and historical information about their evaluations of line pilots could be included. The CASRS would typically include de-identified textual summaries of carrier incidents categorized by fleet. The PKSA database could include demographic data such as pilot experience, (total hours, hours in type), individual skill/ability assessments, and other measures such as attitudes (e.g. FMAQ).

These databases are potentially a rich source of information to optimize training, answer carrier questions and solve operational problems. These databases could form the nucleus of a management information system that would use this information not only to provide the regular monthly reports of baseline data, but also for the statistical flagging of significant operational problems and an in-depth analysis of the causes and possible cures of these operational problems. However, since these DB s are not currently connected and integrated, queries that require information across the databases are difficult or impossible. Typically, questions can be asked for information contained in a single database but NOT questions that require information from multiple databases. This severely limits the questions that can be asked of the data to very simple questions such as questions about fleet differences or trends over time. Limiting the questions that can be asked limits the operational problems that can be solved using this information. The ability to follow up results from initial questions with more detailed analyses exploring reasons for the observed results is lacking. Exploring the reasons for the observed results, particularly results indicating poor or unsafe performance, is critical for designing correct problem solutions or interventions.

Examples of carrier questions that would require integration across databases include: Why did the percent of pilots failing initial qualification increase this year? Why are some of the I/Es more effective instructors than others? What additional training would help I/Es with low effectiveness? How does pilot performance on last

year's recurrent LOE point to necessary instructional curriculum changes? What parts of a pilot's training performance during initial qualification predict continuing qualification performance? When have I/Es drifted off calibration benchmarks enough to require more IRR training? Which individual pilot knowledge, skills, or abilities (KSAs) really predict pilot performance? Which training significantly changes these KSAs? To what extent do different types of CRM training experiences predict later line performance? Answering questions like these requires an integrated repository of carrier database information such as a Data Warehouse.

DATA WAREHOUSE (DW)

Concept

Answering questions like the above requires an integrated, interconnected set of databases. A DW is a set of interconnected DB s that is designed to give maximal information for a broad set of possible questions. More specifically, a DW is a historical collection of point-in-time data about relevant entities (pilots, I/Es, the training program) consolidated from multiple sources. The data are "read-only" (not changed by users) and cover both recent (last month, year) and historic (past years) times. The query interface should be designed to make this information as easily accessible as possible and allow systematic analyses of the data.

Construction

The starting point for constructing a DW is a set of relational databases. Programming a relational database is covered by Ullman (1982). To construct a DW with maximally usable information, the information in different relational databases must be appropriately connected or linked and stored over time. A de-identified PIN numbers can be used to connect pilot information from early stages of a pilot's employment (prior experience, background, hiring evaluation results, fleet common indoctrination training) to later stages (qualification results, continuing qualification results, transition training results). This core of pilot background, training and assessment information must be connected to other databases necessary to answer questions or solve problems such as the Program Audit DB and the Instructor/Evaluator DB. Figure 1 graphically shows the links among these DB s in the DW. The IEDB is linked to pilot training and evaluation information via instructor/evaluator PIN (IPIN) numbers. The PADB is linked to pilot training and evaluation information via systematic content links of curriculum elements and objectives to pilot training (e.g. ground school, LOFT) or testing (e.g. LOE, maneuvers validation) events. For more details on planning and designing a DW, see Barquin and Edelstein (1996).

Routine use

Question development. A DW can answer a wide range of basic carrier questions. Typically, these questions will focus on pilot performance or on the connection of pilot performance to other training and assessment issues. Questions focusing on the pilot performance core of Figure 1 would include:

Is pilot performance improving or declining over time? (Trends)
Does initial qualification performance predict continuing qualification performance?
How do pilot performance results indicate which types of pilots should get extra training? (Outliers)
Can pilots who will have low continuing qualification performance be identified early on in training?
What type of extra training would pilots identified for potential low performance need?

The linked IEDB and PADB would be necessary to answer questions connecting pilot performance to the instructor/evaluators or the curriculum content:

Do some I/Es train pilots to have particularly high performance?
If so, what characterizes these I/Es and how can we duplicate their effectiveness?
How can the performance results be used to modify training by identifying which training topics are
Over trained? (reduce training emphasis)
Under trained? (increase training emphasis)

Which type of training is most effective for increasing pilot performance on LOE? on maneuvers? on line checks?

Analysis plan. The set of basic carrier questions should be connected to a set of routine analyses which answer those questions. These analyses should be organized into an analysis plan which schedules the analyses and reports (e.g. monthly, quarterly, yearly intervals). For each question, the relevant data and appropriate analysis should be clearly stated. For many questions, the appropriate analysis will be simple tabulations, descriptive statistics such as percentages, or simple inferential statistics (e.g. Gravetter and Wallnau, 1992). Statistics are useful to set an objective criterion to trigger reports that require an immediate action. An example is finding a significant decline in pilot performance, which would trigger special reports to appropriate persons (e.g. Fleet manager, Quality Assurance, Training), and require corrective actions. Each report should include graphs (e.g. trends), tables (e.g. percentages), or other forms of presenting information which best communicate the result to the user.

Analysis process. The process of carrying out the data analysis consists of several steps: 1) Stating the question precisely, 2) extracting the relevant data from the DW, 3) checking data types and scoring, 4) combining relevant data into a data table, 5) preparing data for analysis, 6) checking distributions or other analysis assumptions, 7) conducting the analysis, 8) interpreting the analysis to answer the question. The data extraction, combination, preparation, checking, and analysis steps can be largely automated. Checking the type and meaning of data prior to the analysis and interpreting the results of the analysis require human judgment.

Analysis Interpretation. The interpretation of the results depends on the quality of the information on which the analysis is based and the power and precision of the statistical analyses. The adage "Garbage In, Garbage Out" emphasizes that the quality of answer is limited by quality of data. Precise, high-quality answers require sensitive, reliable, and valid data. Obtaining this type of data is facilitated by good design and content of measurement instruments and training of evaluators such as Inter-Rater Reliability training. Statistical power is the ability of a statistical analysis to detect an effect of a certain size. Statistical power depends on the reliability of the data, the amount of data, the size of effect the analysis is trying to detect or estimate, and error rates set for the analysis (Cohen, 1988).

To correctly interpret a result, the user should understand the assumptions and limitations of each analysis technique. Each technique makes assumptions about the data and about the question being asked. Assumptions about the data should be at least plausible for the data being analyzed and, wherever possible, objectively checked. Statistical assumptions about the question being asked should match the real question and be plausible to the user. Different data or questions may require different techniques.

Examples of basic analyses. Basic analyses of DW information will answer many important carrier questions. Figure 2 shows a hypothetical example of detecting performance trends over time. The steady downward drift of Line Check averages and the 2nd Quarter spike upward in LOE averages both require further exploration using information in the Data Warehouse. Figure 3 shows an example of detecting extremely poor pilot performance (outliers). Using statistical methods, the pilots with Line Check averages around 2.25 are significantly below the group average and can be assigned to additional training, more frequent evaluation, or both. Figure 4 illustrates the first step in using the performance database information to adjust training emphasis. Significantly higher or lower performance across LOE items can be linked back to corresponding curriculum content to modify training.

Data Mining

Data Discovery or Data Mining is using data to answer very general or exploratory questions. Questions such as "Why are one fleet's pilots having trouble during Initial Operating Experience?" may require several exploratory analyses which examine possible causes of poor pilot IOE performance. Alternatively, questions such as "What pattern of training performance distinguishes pilots who require repeat training during initial qualification?" require searching the DW for useful patterns or profiles in the data that distinguish a target group of pilots. Data mining techniques include techniques focused on describing the data as well as techniques focused on making inferences about the data or combining inference and description.

Description. Basic data mining techniques include simple descriptive methods such as data graphing (bar charts, graphs, plots, etc.), cluster analysis, and association (e.g. cross-tabulation, contingencies, correlations). More advanced descriptive techniques include factor analysis for finding the underlying dimensions of a set of items, multidimensional scaling to find the descriptive structure underlying a pattern of similarity and dissimilarity among a set of objects, and neural network analysis to find optimal ways of classifying cases using patterns of information.

Inference. Basic data mining techniques also include simple inferential methods such as t-tests, Analysis of Variance (ANOVA), and regression. More advanced inferential techniques include Multivariate ANOVA and multiple regression (e.g. Tabachnick and Fidell, 1989). Advanced techniques which yield both inferential and descriptive information are canonical correlation, discriminant function and classification Analysis, and structural equation modeling (Marcoulides and Schumacker, 1996). An example of an question that could use canonical correlation is "How do the components of initial qualification performance relate to different aspects of continuing qualification performance?".

In general, the more advanced data mining techniques make more assumptions but also offer more complete answers to complex questions. The entire set of basic and advanced analyses offer a wide diversity of powerful techniques for answering a wide variety of carrier questions.

DISCUSSION

DW concept can help integrate required AQP databases with other databases to answer critical carrier questions.
For example, trend analyses on line check performance can "Red Flag" operational problems before the problems become safety critical. Objective triggers can also be set for requiring I/Es to be recalibrated in standardization sessions, altering the training curriculum, and other important issues. These objective triggers can signal the need for management intervention and follow-up analyses to give information and help direct the management decisions.

Beyond monthly reports and triggers, the accumulated information in the DW allows the carrier to answer many relevant questions and solve important problems. Carriers may assess the effectiveness of training and adjust training resources for maximum effectiveness. The antecedents of excellent or poor pilot performance can be tracked back to curriculum content, I/E training, pilot experience, or other relevant causes that determine which interventions are necessary to increase performance. Pilot performances which are poor enough to be classified as extreme outliers can be used to objectively indicate which pilots need special tracking or training.

Data mining techniques can be used to further explore the information in a DW. This exploration can give more complete information for management decisions, and possibly develop new relevant questions. The power of the DW concept lies in combining a rich set of integrated information with a wide variety of potential analytical procedures that can be used to answer carrier questions and test theories of pilot performance. The applicability of performance theories to different types of pilots and types of operational conditions can be competitively assessed.

REFERENCES

Barquin, R.C. and Edelstein, H.A. (Eds.) (1997). Planning and designing the data warehouse. Upper Saddle River, New Jersey: Prentice Hall PTR.

Cohen, J. (1988). Statistical power analysis for the behavioral sciences. Hillsdale, New Jersey: Lawrence Erlbaum Associates, Publishers.

Gravetter, F.J. and Wallnau, L.B. (1992). Statistics for the behavioral sciences. New York, NY: West Publishing Company.

Marcoulides, G.A. and Schumacker, R.E. (Eds.) (1996). Advanced structural equation modeling: Issues and techniques. Mahwah, New Jersey: Lawrence Erlbaum Associates, Publishers.

Tabachnick, B.G. and Fidell, L.S. (1989). Using Multivariate Statistics (2nd Edition). New York, NY: HarperCollins Publishers, Inc.

Ullman, J.D. (1982). Principles of database systems (2nd edition). Rockville, MD: Computer Science Press, Inc.

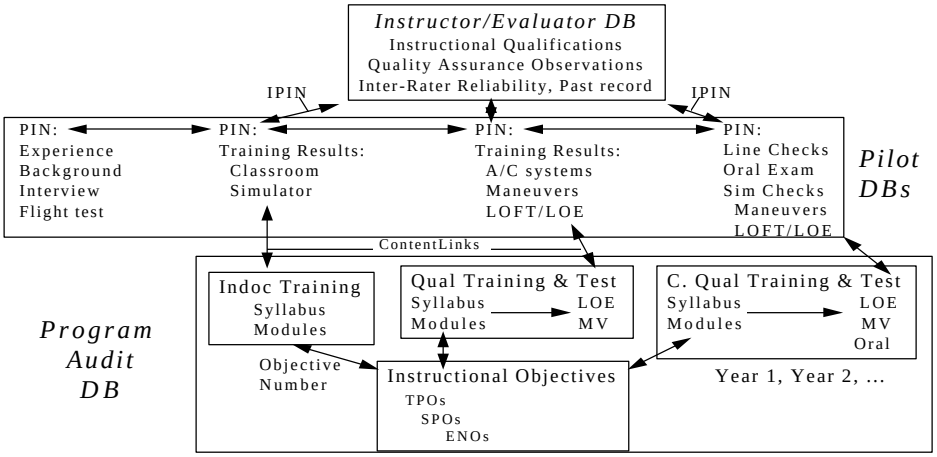

Figure 1. Possible arrangement of AQP databases in a Data Warehouse.

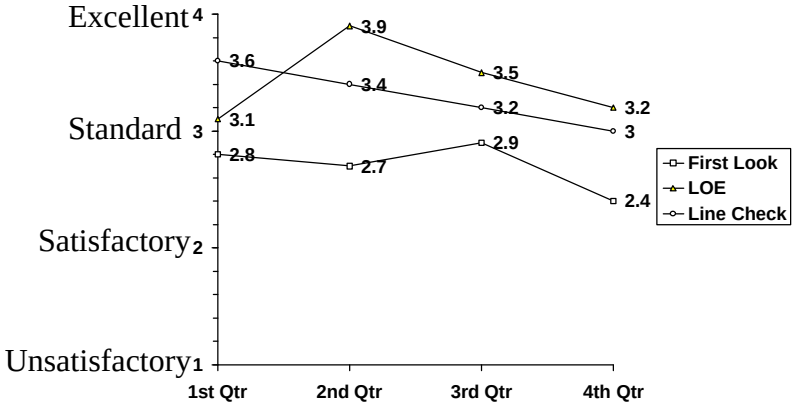

Figure 2. Example of First Look, LOE, and Line Check trends in pilot performance over time (quarters).

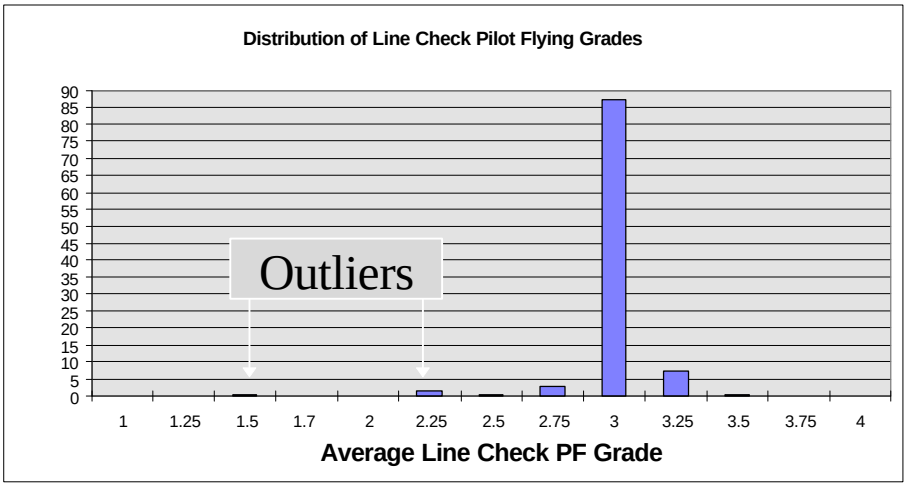

Figure 3. Using statistical outlier detection to objectively classify pilots with poor performance.

2_OBS1

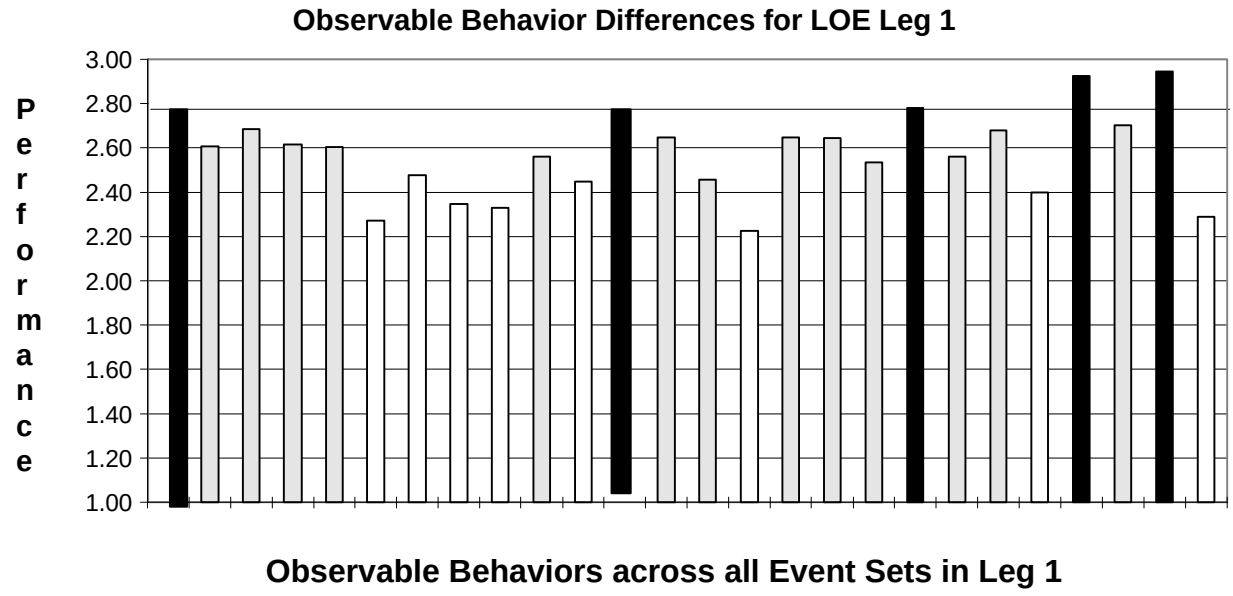

Figure 4. Performance differences on observable behaviors on Leg 1 of an LOE. Significantly low performance behaviors (clear bars) should be linked back to the curriculum areas that require increased training. Significantly high performance (black bars) could be areas for reduced training emphasis.