

Development of Air Traffic Control Measures Database

Esa M. Rantanen and Peter M. Vlach
University of Illinois at Urbana-Champaign

Availability of measures that would predict controller success in his or her task and the impact of changing procedures and advancing technology on the system as a whole is imperative to the success of modernization of air traffic control (ATC) systems worldwide. This paper describes a database that is populated by the results of previous reviews of ATC research literature, organized according to a novel ATC measures taxonomy, and made accessible via the World Wide Web and a purpose-built web interface. The database will also facilitate continual updates, allowing for growth and relevance of its contents into the foreseeable future.

INTRODUCTION

In 1980 V. D. Hopkin's paper titled 'The measurement of the air traffic controller' (Hopkin, 1980) highlighted two critical aspects of measurement in air traffic control (ATC): First, there is a considerable demand for measures for a variety of purposes (see also Manning & Stein, in press, for a recent review), and second, measurement in ATC is often hampered by unique difficulties not found in other domains. Yet, since scientific research and subsequent engineering applications are dependent on methods of measurement (Chapanis, 1959; Meister, 2001), the need for reliable and valid measures is particularly compelling in the research and development (R&D) activities on a variety of aspects of ATC. Ever-increasing traffic volumes and the collateral demands to improve aviation safety and the air transportation system efficiency present an perpetual challenges to the world's ATC system and necessitate the introduction of new technologies and automation applications in ATC. While potentially allowing given airspace systems to accommodate the increasing demand, these technologies will also fundamentally change the system's functionality as well as air traffic controllers' working methods, strategies, workload, and performance (Hopkin, 1995; Wickens, Mavor, & McGee, 1997; Wickens, Mavor, Parasuraman, & McGee, 1998). Thorough assessment and evaluation of the consequences of new technologies on system capacity and safety, as well as the working conditions and performance of individual controllers, is hence of utmost importance. Success of these evaluation efforts, however, is subject to the availability of valid and reliable measures.

The subsequent 27 years since publication of Hopkin's paper have seen many major developments in ATC: In 1981 the Federal Aviation Administration (FAA) introduced the National Airspace System (NAS) Plan. This plan included the Automated En Route Air Traffic Control (AERA) project and the Advance Automation System (AAS) Common Console, which were to revolutionize ATC and reassign many tasks and functions performed by human controllers to computers. The AERA and AAS systems were very ambitious and technical problems resulted in cost and schedule overruns and eventual termination of the program. Work for NAS modernization

continued, however. In 1995, the RTCA (formerly the Radio Technical Commission for Aeronautics) published reports on the concept of free flight (RTCA, 1995a, 1995b), and while the technological advancements to allow for implementation of this concept have been evolutionary rather than revolutionary in nature, the overall R&D efforts in ATC continue unabated. Today, many technologies that ultimately will make mature free flight possible are already being implemented in operational use. Similar developments are under way in Europe and Australia, and there is no reason to think that the pace of modernization of ATC systems worldwide would slow down. It is against this background that measurement issues in ATC must presently be examined.

The intricacies of measurement in ATC have, if possible, only been exacerbated by the advent of automation applications throughout the domain, increasing the complexity of the task environments in which controllers work. Consequently, scientific investigation regarding the impact of new technologies has become increasingly difficult due to the escalating number of variables and their interactions in the operational environments, as well as the shift from overt performance (i.e. manual control) to predominantly covert behavior (i.e. supervisory control) of the operators. It may also be argued that the traditional measurement techniques in ATC, that is, subjective, over-the-shoulder (OTS) evaluation of controller performance by other experienced controllers, have become inadequate in the face of the present challenges. While sound and valid arguments have been made both for (e.g., Hennessy, 1990) and against (e.g., Kosso, 1989; Scheffler, 1967) the use of subjective measurements, it is clear that objective measures are highly desirable. Such methods could be used both in conjunction with high fidelity, realistic ATC simulation and in operational settings. Additionally, objective measures can be routinely collected and analyzed concurrently and in an unintrusive manner during the task, and subjected to data mining techniques to detect trends in the system's performance, before any possible problems are manifested as incidents or operational errors.

The availability of measures in ATC, as in any scientific endeavor, is largely contingent upon the availability of data in a usable form. Collection of usable data from an environment

as dynamic, variable, and multidimensional as ATC is not a trivial matter. ATC modernization efforts, as well as recent advances in the area of digital technology, however, potentially offer new sources for data, along with data collection and storage methods. An example of access to data from which ATC measures can be derived is the System Activity Recordings (SAR) that stores all flight and radar information in Air Route Traffic Control Centers (ARTCCs). These data can be further processed the Data Analysis and Reduction Tool (DART) (Federal Aviation Administration [FAA], 1993) and the National Track Analysis Program (NTAP) (FAA, 1991), which produce a number of text-based output files that can be further analyzed by specialized applications, such as the Performance and Objective Workload Evaluation Research (POWER) software (Mills, Manning, & Pfleiderer, 1999; Manning, Mills, Fox, & Pfleiderer, 2001), which currently derives over 40 separate measures, describing a variety of aspects of ATC.

Purpose of the Research

In summary, vast amounts of objective data are potentially available from the operational ATC system. The locus of the problem is therefore not in availability of data, but rather in derivation of valid, reliable, and meaningful measures from the abundance of data. This problem is particularly pronounced when variables of interest are not directly measurable, such as controller workload, situation awareness, and performance.

Hopkin (1980) concluded his paper by contemplating 'prospects for progress' (p. 555) and called for development of better measures of the air traffic controller. Much research on a wide variety of ATC-related issues has since been conducted. Hadley, Guttman, and Stringer (1999) listed no less than 170 separate measures or measurement techniques in their air traffic control specialist (ATCS) performance measurement database. Many of these measures, however, have substantial overlap, are derivatives of each other, and measure diverse aspects of ATC functions, all of which are not relevant to the performance of an individual air traffic controller.

Our work focused on organizing measures reported in ATC research literature in some meaningful way that would allow for evaluation of the present state of the art of ATC research, gauge the progress made since Hopkin's (1980) paper, and identify areas that are ripe for further research. To accomplish these objectives, we developed a taxonomy of measures that will allow for cross-referencing between different types of measures, their purposes, and the required data, potentially facilitating development of comprehensive models of ATC performance and additional measures as new sources of data become available.

TAXONOMY OF ATC MEASURES

The taxonomy presented here is based on an extensive review of literature related to ATC R&D. We catalogued and classified measures used in published research, and hence the lowest level of classes in our taxonomy is based on existing research in this domain. We predominantly sought out original

and empirical papers, but in a limited number of instances had to rely on secondary sources, that is, on review articles. The most notable exception to this standard is the Hadley et al. (1999) ATCS performance measurement database, which was included in our review in its entirety.

Direct Measures

The main division of measures in the taxonomy presented here is between direct and indirect measures. Direct measures are defined as those that can be explicitly measured, for example, by direct observation of a controller's action, measurement of a response latency, or count of aircraft in a sector at a given time. A direct measure is thus a number directly associated with a physical quantity or natural phenomenon and compared to a standard, that is, a unit. Direct measures are expressed in terms of standard units of time (milliseconds, seconds, minutes, etc.), length (millimeters, inches, nautical miles, etc.) and its derivatives area and volume, as well as units of mass, electrical current, force, etc., and count. For example, the NASA-TLX index asks subjects to rate their workload along several scales, including one titled 'Mental Load'. A direct measure in this case is the numerical rating the subject provides (e.g., '4'); this rating then serves as an indirect measure or mental load.

Identification and classification of direct measures in the literature was a relatively simple task. In the resulting taxonomy, the separately reported direct measures were grouped under 65 distinct classes, down to sixth level in some cases. The taxonomic structure is based on published research literature and the level of classification was dependent on the details provided in the literature.

Direct Subjective and Direct Objective Measures

Within the class of direct measures, the next sublevel was created by differentiating between subjective and objective measures. The method for making this distinction is identification of objective criteria, a prerequisite for meaningful measurement (Meister, 1989). The classification is here based on an objective criterion against which the observer bases his or her judgment. For example, an FAA OTS evaluation sheet (Form 3430) contains items such as 'standard phraseology is not adhered to' and 'awareness is not maintained.' The former has an objective criterion, the standard phraseology as published in the FAA Air Traffic Control Handbook (7110.65) and hence any deviations from it would warrant a check in the aforementioned box. However, the latter, despite guidelines for making this judgment, provides much latitude for a subjective assessment.

The subcategory of direct, objective measures was further divided in two major sub-sub categories, system measures and human measures. Variables that were not directly dependent on human performance were classified as system measures. Hence, for example, number of keystrokes required to enter data into the system is system-related; there is little the controller can do about it, although it can certainly be used as a possible metric of task load (an indirect measure). Keystroke errors, on the other hand, would be classified as a human

measure. Direct, objective, human measures fall under further three categories: observer-rated (recall the role of explicit criteria in this class), cognitive, which is further divided into two time-based classes, response time and time required to perform a task, and a variety of psychophysical metrics.

Objective observer-rated measures predominantly consist of either graded exam-type questionnaires or problems, to which a correct answer can be known, or observed behavior or performance compared to a given standard. Response time is extensively used to probe a wide variety of covert cognitive processes, and it is also readily available and measurable in many settings. Time required to perform a task as a separate category is justified by a similar class of system measures and as a counterpart of the time required vs. time available paradigm, which is a basis of some task load and complexity constructs (Chatterji, & Sridhar, 2001; Chiles, & Alluisi, 1979; Mogford & al. 1995). Very few cases where physiological measures were used in the ATC research were found. The vast majority of physiological measures were different eye movement measures.

Indirect Measures

An indirect measure measures a variable that cannot be directly measured, but which may be inferred based on numerical values of directly measured variables. For example, certain actions of a controller may be indicative of his or her performance, response latency can be used to make inferences on some covert cognitive processes, and a number of aircraft in a sector can be used to signify sector complexity. Indirect measures were classified in a similar manner as direct measures under 36 distinct categories, based on the reviewed literature.

The main sub-classification scheme under indirect measures was between human and system measures. Of human measures, workload, situation awareness, and performance

were of most interest. Of system measures, airspace or sector complexity has attracted much attention in recent years (Kopardekar & Magyarits, 2002). The definitions of the indirect constructs varied widely, however, and classification of indirect variables is therefore based only on authors' own account of the measure.

DATABASE DESIGN

Basic Database Structure

The ATC measures database was constructed using the MySQL open source, relational database management system. The database was structured to consist of three tables: The direct taxonomy table and indirect taxonomy table are both used to provide classification and search of the literature by measure categories; the literature table holds the desired information within the scope of this database. Each table can be accessed via a specific web interface for adding classes in the taxonomy tables or editing existing classes, and in the case of the literature table, to enter new data into the database

The Taxonomy tables are linked to the Literature table via their corresponding rec_ids, t1_rec_id for direct taxonomy and t2_rec_id for indirect taxonomy. This design provided for necessary means for easy and independent data entry in each taxonomy table and will facilitate future changes in the taxonomies. These values are used in the literature table in form of pull-down menus for easy addition of literature citations to the database and linking these entries to the measure taxonomies.

Each entry in the literature table has a link to its corresponding taxonomy item. Since the literature table includes links to both taxonomy tables, the system can be used to cross-reference direct and indirect taxonomy every time the literature record is retrieved. Furthermore, the database can be searched independently by direct or indirect measures. See Figure 1 for the database structure.

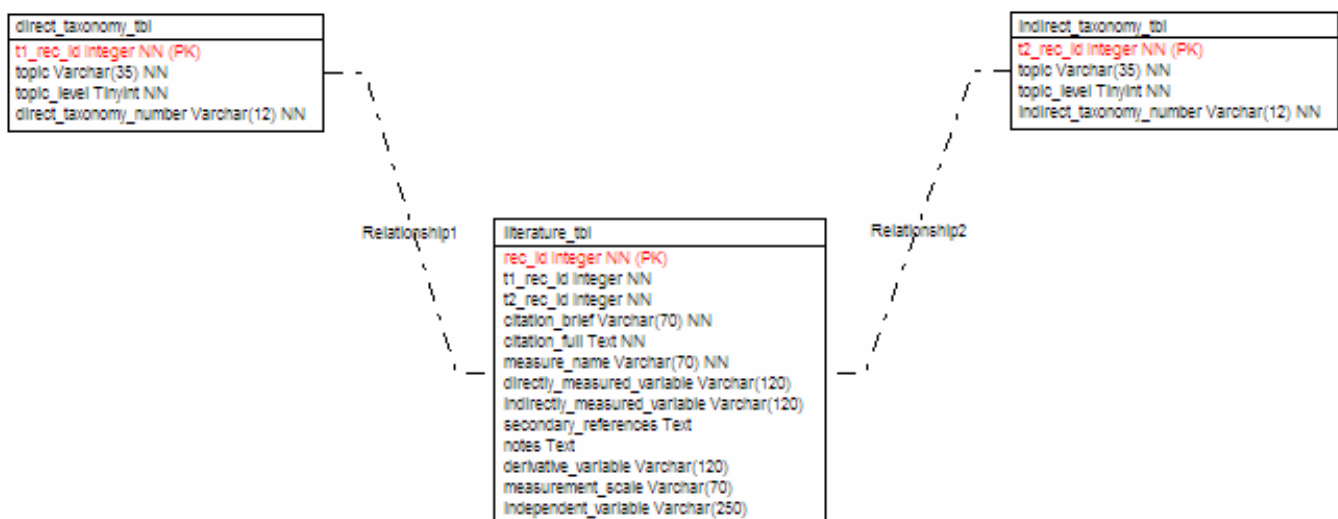


Figure 1. The three tables and their relationships in the ATC measures database. The direct and indirect Taxonomy tables are both used to provide classification and search of the literature by measure categories; the literature table holds the desired information within the scope of this database.

Other Attributes of ATC Measures

In addition to the ATC measures classification into direct and indirect measure and their respective subclasses described above, several other attributes of measures were also included in the database. These include the following:

Measure name: This is the name of the measure as used by the authors of the source verbatim, if a specific name of the measure was used. Examples of such names (or acronyms) are NASA-TLX, SAGAT, SPAM, etc.

Measurement scale: This indicates the scale of the measurement used in the article entered into the database. There are four levels of measurement scales that are commonly distinguished: nominal, ordinal, interval, and ratio scales.

Directly and indirectly measured variables: In addition to the direct and indirect measures included in the taxonomy, the database allows for recording the name or names of the variables verbatim as described in the original article. Directly measured variables are expressed in terms of standard units of time length, etc., but may be more specific than those included in the taxonomy; of course, having an identical entry in both ‘Measure’ and ‘Measured Variable’ fields is certainly very possible

Independent variable: We also recorded the name or names of the independent variables as described in the article (verbatim) in the database. As these variables are manipulated in an experiment or study and whose presence or degree determines the change in the dependent variable, which is measured, inclusion of independent variables in the database will allow for mapping of particular measures to given topics of interest.

Derivative variable: This refers to the statistical transformations the directly measured variables were subjected to in reporting of the results, for example measures of central tendency (mean, median and mode) or dispersion (variance or standard deviation).

Secondary references: Finally, we included in the database references the author(s) of the article may have used to make the connection between the direct and indirect variables in their research. This information will potentially allow for tracking the sources of given measures and their development history, although it is dependent on the completeness of references in the articles where the measures were extracted.

DATABASE USERS’ GUIDE

The database (literature table) has been initially populated with the citations from literature review by Rantanen (2004). A web search interface allows for searching the database by direct or indirect measures. The searcher may choose any level in the taxonomy for return of everything under that level, including literature under any sublevels. The search term choice is done simply by a pull-down menu containing the measures taxonomy (see Figure 2). The output includes a list of relevant literature, with a name of the measure and its taxonomy number, as well as the indirect measure pertaining to the research (or, direct measure, if the database was searched by indirect measure). Clicking on the brief citation field will show a complete record in a new window.

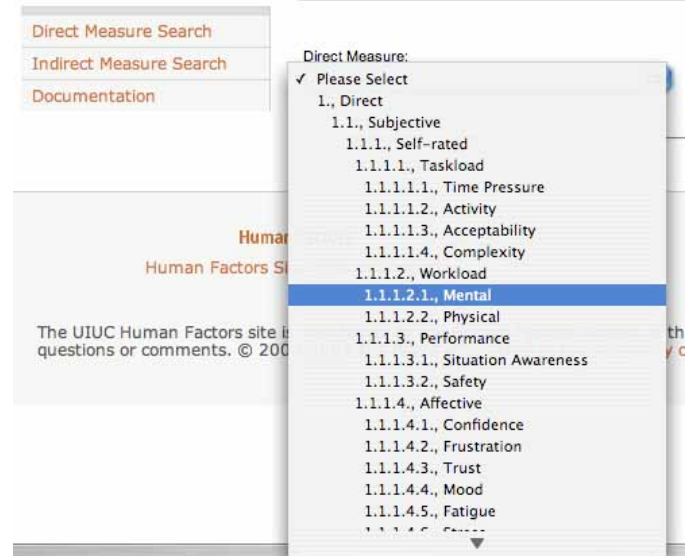


Figure 2. A (partial) screenshot of the direct measure search interface. A search term (direct measure) is simply selected from a pull-down menu containing the taxonomy.



Figure 3. The data entry interface for the literature table. The direct and indirect measure entries will be done through the same pull-down menus as on the search interface (see Fig. 2). If an article contains multiple metrics, each will be entered separately; a ‘copy’ function makes multiple entries easy.

A literature database such as this will naturally require continual updating to remain valid and relevant. The literature citations entry screen can be made available to any number of contributors, who wish to commit to entering new studies and ATC measures into the database. The literature citation entry screen (Figure 3) and the structure of the measurement taxonomy are straightforward enough to allow anybody to contribute to the database with minimal training and practice.

Since the measures entered into the database must be selected from the taxonomy in the pull-down menu, and since the measures taxonomy is based on past research—albeit an exhaustive review of relevant literature and the individual classes are meant to be broad and general—it is conceivable that novel metrics are developed that do not easily fall within the current taxonomy. For this reason, the literature table also contains a ‘notes’ field, where such instances may be recorded for periodic reviews. The taxonomy may then be updated as warranted.

SUMMARY AND CONCLUSION

This project was undertaken to create a database of different measures and metrics used in ATC R&D, as reported in published literature, and to make it usable for a wide variety of individuals interested or involved in ATC related research. The resultant database and the associated interfaces and functionalities must at this time be considered a beta version. No systematic usability studies on it have yet been performed. As the database and its maintenance tools are published on the web, it is hoped that a number of researchers in the ATC community would choose to contribute to the continual updating of the database as new research is published or old, relevant, but perhaps obscure, studies are found. Through these collaborative efforts the database can be built into a truly useful tool for all future ATC R&D efforts worldwide.

ACKNOWLEDGMENTS

This material is based upon work supported by the Federal Aviation Administration under Cooperative Agreement No. DOT 02-G-019 and Purchase Order # DTFAAC-06-P-08010. Drs. Carol Manning and Jerry Crutchfield were the scientific/technical monitors, respectively. Any opinions, findings, conclusions, and recommendations expressed here are those of the authors and do not necessarily reflect the views of the Federal Aviation Administration.

REFERENCES

- Chapanis, A. (1959). *Research techniques in human factors*. Baltimore, MD: Johns Hopkins University Press.
- Chatterji, G., & Sridhar, B. (2001). Measures of air traffic controller workload prediction. *First AIAA Aircraft Technology, Integration, and Operations Forum* (AIAA Paper 2001-5242). Los Angeles, CA, Oct. 16-18, 2001.
- Chiles, W. D., & Alluisi, E. A. (1979). On the specification of operator or occupational workload with performance-measurement methods. *Human Factors*, 21(5), 515-528.
- Federal Aviation Administration (1991). *Multiple Virtual Storage (MVS); Subprogram Design Document; National Track Analysis Program (NTAP)*. NASP-9247-PO2. Washington, DC: Author.
- Federal Aviation Administration (1993). *Multiple Virtual Storage (MVS); User's Manual; Data Analysis and Reduction Tool (DART)*. NASP-9114-H04. Washington, DC: Author.
- Hadley, G. A., Guttman, J. A., & Stringer, P. G. (1999). *Air traffic control specialist performance measurement database*. (DOT/FAA/CT-TN99/77). Atlantic City, NJ: FAA
- Hennessy, R. T. (1990). Practical human performance testing and evaluation. In H. R. Booher (Ed.), *MANPRINT: An approach to systems integration* (pp. 433-479). New York: Van Nostrand Reinhold.
- Hopkin, V. D. (1980). The measurement of the air traffic controller. *Human Factors*, 22(5), 547-560.
- Hopkin, V. D. (1995). *Human factors in air traffic control*. Bristol, PA: Taylor & Francis.
- Kopardekar, P., & Magyarits, S. (2002). Dynamic density: Measuring and predicting sector complexity. *Paper presented at the 21st Digital Avionics Systems Conference*, Irvine, CA.
- Kosso, P. (1989). Science and objectivity. *Journal of Philosophy*, 86, 245-257.
- Manning, C. A., Mills, S. H., Fox, C. M., & Pfeleiderer, E. (2001). *Investigating the validity of performance and objective workload evaluation research (POWER)*. DOT/FAA/AM-01/10. Washington, DC: Office of Aerospace Medicine.
- Manning, C., & Stein, E. (In press). Measuring air traffic controller performance in the 21st century. In B. Kirwan, M. Rodgers, & D. Schaefer (Eds.), *Human Factors Impacts in Air Traffic Management* (Chapter X). Ashgate Publishing Company.
- Meister, D. (1989). *Conceptual aspects of human factors*. Baltimore, MD: The Johns Hopkins University Press.
- Meister, D. (2001). Basic premises and principles of human factors measurement. *Theoretical Issues in Ergonomics Science*, 2(1), 1-22.
- Mills, S. H., Manning, C., & Pfeleiderer, E. M. (1999). Computing en route baseline measures with POWER. *Poster presented at the 10th International Symposium on Aviation Psychology*, Columbus, OH, May 3-6, 1999.
- Mogford, R. H., Guttman, J. A., Morrow, S. L., & Kopardekar, P. (1995). *The complexity construct in air traffic control: Review and synthesis of the literature*. DOT/FAA/CT-TN95/22. Washington, DC: FAA, Office of Aviation Research.
- Rantanen, E. M. (2004). *Development and Validation of Objective Performance and Workload Measures in Air Traffic Control* (AHFD-04-19/FAA-04-7). Savoy, IL: Av. Human Factors Div.
- RTCA (1995a). *Report of the RTCA Board of Directors' Select Committee on Free Flight*. Washington, DC: Author.
- RTCA (1995b). *Final Report of RTCA Task Force 3 Free Flight Implementation*. Washington, DC: Author.
- Scheffler, I. (1967). *Science and subjectivity*. Indianapolis, IN: Bobbs-Merrill.
- Wickens, C. D., Mavor, M., & McGee, J. (1997). *Flight to the future: Human factors of air traffic control*. Washington, DC: National Academy of Science.
- Wickens, C. D., Mavor, M., Parasuraman, R., & McGee, J. (1998). *The future of air traffic control*. Washington, DC: National Academy of Science.