ote technical note techn

Tower Cab Metrics

Randy L. Sollenberger, Ph.D.
Pamela S. Della Rocco, Ph.D.

DOT/FAA/CT-TN02/03
December 2001

U.S. Department of Transportation
Federal Aviation Administration

William J. Hughes Technical Center
Atlantic City International Airport, NJ 08405

Technical Report Documentation Page

1. Report No. DOT/FAA/CT-TN02/03	**2. Government Accession No.**	**3. Recipient's Catalog No.**

4. Title and Subtitle Tower Cab Metrics	**5. Report Date** December 2001
	6. Performing Organization Code ACT-530
7. Author(s) Randy L. Sollenberger, Ph.D., and Pamela S. Della Rocco, Ph.D.	**8. Performing Organization Report No.** DOT/FAA/CT-TN02/03
9. Performing Organization Name and Address Federal Aviation Administration William J. Hughes Technical Center Atlantic City International Airport, NJ 08405	**10. Work Unit No. (TRAIS)**
	11. Contract or Grant No.
12. Sponsoring Agency Name and Address Federal Aviation Administration Human Factors Division 800 Independence Ave., S.W. Washington, DC 20591	**13. Type of Report and Period Covered** Technical Note
	14. Sponsoring Agency Code AAR-100

15. Supplementary Notes

16. Abstract

This report is part of a continuing effort to develop human factors measures for different operational environments in the Federal Aviation Administration Air Traffic Control (ATC) system. Previous research at the William J. Hughes Technical Center Research Development and Human Factors Laboratory has focused on measures for the Terminal Radar Approach Control and en route radar environments (e.g., Sollenberger, Stein, & Gromelski, 1997; Vardaman & Stein, 1998). This work focuses on human factors measures for the local and ground controller positions in the Tower Cab environment. We reviewed a number of existing sources of measures that have been used in different ATC environments. The result is a proposed battery of objective and subjective candidate measures for the Tower Cab. In addition, we present an initial effort to develop behavioral rating scales for use by subject matter experts to evaluate performance in the Tower Cab environment. As a whole, these measures provide a broad look at performance for researchers to apply in a simulation environment. This allows for conducting experiments to obtain both baseline and comparison measures for assessing the effects of any proposed changes in technology, automation, or procedures. We must conduct more research to assess the validity and reliability of the proposed measures in the Tower Cab environment through continued usage in simulations. Future research should also develop metrics for the remaining positions in the Tower Cab.

17. Key Words Air Traffic Control, Human Factors, Performance, Measures, Metrics, Rating Scales, Tower Cab	**18. Distribution Statement** This report is approved for public release and is on file at the William J. Hughes Technical Center, Aviation Security Research and Development Library, Atlantic City International Airport, New Jersey 08405. This document is available to the public through the National Technical Information Service.

19. Security Classif. (of this report) Unclassified	**20. Security Classif. (of this page)** Unclassified	**21. No. of Pages** 29	**22. Price**

Form DOT F 1700.7 (8-72) **Reproduction of completed page authorized**

Table of Contents

Appendixes

List of Illustrations

Executive Summary

This technical report is part of a continuing effort to develop human factors measures for different operational environments in the Federal Aviation Administration Air Traffic Control (ATC) system. Previous research at the William J. Hughes Technical Center Research Development and Human Factors Laboratory has focused on measures for the Terminal Radar Approach Control and en route radar environments (e.g., Sollenberger, Stein, & Gromelski, 1997; Vardaman & Stein, 1998). This work focuses on human factors measures for the local and ground controller positions in the Tower Cab environment.

In this report, we reviewed several existing sources of measures that have been used in different ATC environments. The result is a proposed battery of objective and subjective candidate measures for the Tower Cab. In addition, we present an initial effort to develop behavioral rating scales for use by subject matter experts to evaluate performance in the Tower Cab environment. As a whole, these measures provide a broad look at performance for researchers to apply in a simulation environment. This allows for conducting experiments to obtain baseline and comparison measures to assess the effects of any proposed changes in technology, automation, or procedures.

We intend this battery of measures to serve as a general tool to evaluate performance in the Tower Cab environment. In addition, we recommend that researchers include other measures that are specific to the ideas being evaluated in the simulation. Finally, we must conduct more research to assess the validity and reliability of the proposed measures in the Tower Cab environment through continued usage in simulations. Future research should also develop metrics for the remaining positions in the Tower Cab.

1. Introduction

This research is part of a continuing effort to develop human factors measures for different operational environments in the Federal Aviation Administration (FAA) Air Traffic Control (ATC) system. Previous research has focused on measures for the Terminal Radar Approach Control (TRACON) and en route radar environments (e.g., Sollenberger, Stein, & Gromelski, 1997; Vardaman & Stein, 1998). This work focuses on the Tower Cab.

The purpose of the Tower Cab Metrics Project is to propose a set of human factors measures that could be used in real-time simulation. Although there are several controller positions in an Air Traffic Control Tower (ATCT), this project will focus on proposing measures for the local and ground controllers who directly control aircraft movement on the ground and in the immediate airport vicinity.

In order to propose a battery of measures, we evaluated a number of existing sources. It was our goal to provide both objective and subjective candidate metrics to advance our measurement capability for the Tower Cab environment to the same level as our previous research in the TRACON and en route environments at the Federal Aviation Administration (FAA) William J. Hughes Technical Center (WJHTC) Research Development and Human Factors Laboratory (RDHFL). This report presents the results of our review of existing sources, as well as an initial effort to develop behavioral ratings scales for the Tower Cab.

1.1 Background

The FAA WJHTC is responsible for the development and evaluation of new ATC systems, procedures, and concepts. Much of the research is accomplished by conducting real-time simulations with Air Traffic Control Specialists (ATCSs) where proposed changes to the system are compared to defined requirements or the existing system. As new ATC technologies become available, it is critical to understand the effect the technology will have on ATCS job performance. A critical component of the development and evaluation process is a human factors approach that emphasizes the effectiveness of ATCSs while using the new systems, procedures, or advanced concepts. In order to support these research activities, human factors specialists in the RDHFL have been developing measures that can assess controller performance and overall system effectiveness. To date, little research has addressed the development of performance measures in the Tower Cab environment.

1.1.1 Categories of Air Traffic Control Terminals

The FAA categorizes ATC terminals according to the control services provided. The Position Classification Standard for Air Traffic Control Series ATC – 2152 (FAA, 1998b) describes six types of ATC.

Tower without Radar – Controllers at this category of terminal primarily handle aircraft operating under Visual Flight Rules (VFR). These terminals are located at airports where the principal users are low performance aircraft. These controllers are responsible for air traffic at or in the immediate vicinity of the airport on which the tower is located.

Combination Non-Radar Approach Control and Tower without Radar – In addition to handling aircraft operating under VFR, controllers at this category of terminal also provide (without the use of radar) approach and departure control services to aircraft operating under Instrument Flight Rules (IFR). This type of terminal is located at airports with users similar to those at towers without radar. In addition to providing air traffic control for the airport at which the tower is located, non-radar approach controllers frequently control air traffic operating to and from one or more adjacent airports.

Tower with Radar – Controllers at this category of terminal provide air traffic advisories, spacing, sequencing, and separation services to VFR and IFR aircraft operating within the vicinity of the airport. They use a combination of radar and direct observation. This category includes the single largest number of terminal facilities, ranging from general aviation airports to metropolitan airports serving a wide variety of users. This includes all types of tower facilities that have radar except those that have delegated airspace directly from the parent center.

Combination Radar Approach Control and Tower with Radar – Controllers at this category of terminal provide radar control to aircraft arriving or departing the primary airport and adjacent airports and to aircraft transiting the terminal's airspace. It serves major air carrier airports as well as airports with significant amounts of other users. In contrast to other categories of terminals, combination radar approach controls and towers with radar are divided into two functional areas: radar approach controls and towers. These areas are located within the same facility or in close proximity to one another, and controllers rotate between both areas.

TRACON – Controllers at this category of terminal perform the same radar approach control duties as controllers who work in a combination radar approach control and tower with radar with one major exception – they do not rotate through the tower. In some cases, these facilities are located in the same building as a tower, and, at other locations, they are not. This category of terminal provides radar approach control services into and out of one or more airports.

Combined TRACON Facility – Controllers at this category of terminal provide radar approach control services for two or more large hub airports as well as other satellite airports, where no single airport operations accounts for more than 60% of the total combined TRACON facility air traffic count. A large hub airport is defined as an airport with an annual total facility air traffic count of 300,000 or more. Once designated as a large hub airport, its annual total facility air traffic count may fall as far as 275,000, and it will retain its designation as a large hub airport. If its annual total facility air traffic count falls below 275,000, and it is not projected to increase back to that level, it will lose its designation as a large hub airport. This category of terminal requires such a large number of radar control positions that it must have three or more areas of specialization. Controllers rotate through all positions of operation within each area of specialization. Two or more radar systems are required to provide the ATCSs the surveillance they need to control air traffic throughout the larger operational areas characteristic of combined TRACON facilities.

Each of these facility classifications requires application of different knowledge, skills, and abilities (KSAs). The KSAs are defined in the FAA Position Classification Standard for Air Traffic Control Series ATC – 2152 (1998b).

1.1.2 Tower Positions and Responsibilities

Tower controllers are responsible for control of traffic on the ground, from the ground to departure, and from just prior to arrival to the ground (Wickens, Mavor, & McGee, 1997). They issue the initial clearances, confirming or modifying the filed flight plan. They manage the traffic on the airport surface areas and provide takeoff and landing clearances. They coordinate their actions with both TRACON facilities as well as Air Route Traffic Control Centers (ARTCCs).

To accomplish these responsibilities, there are several distinct operational positions in the Tower Cab. The operating positions are identified in Order 7110.23B (FAA, 1997). The positions include: 1) Tower Operational Supervisor, 2) Tower Cab Coordinator, 3) Clearance Delivery, 4) Flight Data, 5) Ground Control, 6) Local Control and, 7) Traffic Management Coordinator. The positions a tower will have in operation at any given time depend on the traffic levels of the tower and local staffing requirements. During slow periods, these positions may be combined and staffed with a single ATCS, or, during busy periods, one or more Assistant Ground Control and Assistant Local Control positions may be assigned.

The Tower Operational Supervisor directs the operations in the tower on an assigned watch. The Traffic Management Coordinator serves as the point of contact between various organizations and individuals (e.g., the ATC Systems Command Center, Terminal facilities, Operational Supervisors, and ATCSs) to ensure that traffic management procedures, programs, and initiatives are implemented. The Tower Cab Coordinator is responsible for maintaining awareness of the traffic situation to assist in the coordination of facility utilization (e.g., runway utilization, arrivals, and departures) issues with the Traffic Management and TRACON Coordinators.

The Tower Flight Data position ensures the accuracy of all IFR clearances, amends them as necessary, and prepares the Automatic Terminal Information Service (ATIS) information. Clearance Delivery issues the clearances. Clearance Delivery and Flight Data positions may be combined in different facilities.

The Ground Control position controls all aircraft and vehicle traffic in the airport movement area. The Local Control position assumes separation responsibilities for aircraft in a certain radius around the airport from specified altitudes to the ground and on the runways. Local Control handles arrivals and departure sequencing and separation. At smaller airports, these positions may be combined (Wickens et al., 1997). At larger airports, there may be more than one ground and local controller and assistants may be provided for either position. This report will focus on Ground Control and Local Control because these two positions are directly responsible for the control of traffic in the Tower Cab environment.

1.1.3 Job Task Analyses

In order to develop any human factors measures, it is critical to understand the tasks performed, in this case, by local and ground controllers. In 1989, CTA conducted a job task analysis of the Tower positions that we used as part of developing measures of job performance (Alexander et al., 1989). CTA identified 340 tasks for Local Control and rated 187 (55%) as having high or extreme criticality. Tasks for the Local Control position fell into the broad categories of 1) perform local situation monitoring, 2) resolve conflict situations, 3) manage air traffic sequences, 4) route or plan flights, 5) assess weather impact, 6) manage local controller position resources and, 7) respond to system/equipment degradation.

CTA identified 214 tasks for Ground Control and rated 68 (32%) as having high criticality and 122 (57%) as medium criticality (Alexander et al., 1989). Tasks for the Ground Control position fell into the broad categories of 1) perform ground situation monitoring, 2) control aircraft/vehicle ground movement, 3) route or plan flights, 4) assess weather impact, 5) manage ground controller position resources, and 6) respond to system/equipment degradation.

The CTA task analysis identified observable behaviors that can assist in the development of performance measures for the Tower Cab environment (Alexander et al., 1989). Many of these observable behaviors are included in the behaviorally based rating form that we proposed for the Tower Cab environment and will be described later in this paper. This task analysis is just over a decade old, but it provided a valuable resource for developing job performance measures. By specifying job tasks in such detail with frequency and criticality information, it provides a foundation for tracking changes to critical tasks with the introduction of new technologies or procedural changes. The task analysis, however, may be dated. Although it is beyond the scope of this report, we recommend that an updated task analysis be completed.

1.1.4 Previous Research

Researchers at the WJHTC have been conducting ATC simulations and other aviation human factors research since the FAA first established the facility at the Atlantic City International Airport in 1958. Stein and Buckley (1994) compiled a comprehensive bibliography of publications of this research to promote awareness and benefit researchers working in aviation human factors. The bibliography includes not only citations for simulation research but also operational tests and evaluations, workshops, interviews, and survey publications. It also includes an indexing system that categorizes the publications into areas of human factors interest (e.g., visual scanning, workload, and situation awareness).

Buckley, DeBaryshe, Hitchner, and Kohn (1983) conducted a series of real-time ATC simulation experiments using a large set of commonly used simulation measures they called System Effectiveness Measures (SEMs). Examples include operational errors and deviations, runway incursions, and number of aircraft handled. Although the study focused on the en route environment, most of the SEMs were generally applicable to both the en route and terminal radar environments. The SEMs used in the study were based upon the ATC mission goals of safe and expeditious movement of air traffic. The purpose of this study was to investigate the reliability of these measures and applicability of inferential statistics using these measures for evaluating proposed changes to the ATC system. The researchers found that there was large variability in the SEMs, but with large enough sample sizes and adequate training in the simulation environment, inferential statistics could provide reliable results.

Stein and Buckley (1992) selected a subset of the SEMs and provided a specification for measures that should be included in real-time ATC simulation research. The FAA WJHTC RDHFL has developed software for the data collection, reduction, and analysis of these measures and has been using these measures in their ATC simulations since 1993. Most of these measures are general enough to apply to the TRACON and en route environments as well as the Tower Cab environment.

Hadley, Guttman, and Stringer (1999) compiled a database of performance measures that have been used in ATC simulations and human factors research. The database includes a large list of research measures and techniques, their definitions and descriptions, and complete references of published studies that have used the measures. The database consists of four categories of ATC measures: 1) ATCS performance measures, 2) ATC system effectiveness measures, 3) ATCS functional performance model measures, and 4) ATC environment measures. The measures are identified by ATC environment. The database is available in spreadsheet format and users can search by keywords, specific references, or measurement type. Many of the measures in this database are applicable to the Tower Cab environment and have been proposed as candidate measures.

Human factors researchers at the RDHFL have also developed behavioral rating forms to evaluate controller and system performance in real-time simulations in the TRACON and en route radar environments (Sollenberger et al., 1997; Vardaman & Stein, 1998). The researchers designed the observer rating forms for use in research settings by ATCSs experienced in training and evaluating controllers at actual ATC facilities. The rating forms have similar formats that provide a comprehensive assessment of controller performance organized into six major performance categories (or major rating scales) with several specific areas of performance (or minor scales) under each major category. The TRACON form consists of 24 individual rating scales, and the en route form consists of 26 individual rating scales.

The rating forms include a list of observable actions under each rating scale that trained ATCSs can identify to make behaviorally based ratings of controller performance. ATCSs use the list of observable behaviors to improve the reliability of their ratings. The rating forms consist of an 8-point rating system for all rating scales with general performance descriptions associated with each numeric point. The researchers decided that an 8-point rating system was sufficient to ensure rating sensitivity and still allowed observers to discriminate the differences in performance that are associated with each numeric point. The rating forms also include an observation section for each rating scale. ATCSs use the observations sections to describe effective and ineffective controller actions observed during simulation. The observations serve to document the basis of the assigned ratings and help the researchers understand factors considered.

From this body of research, a model for investigation of ATC human factors has evolved. The model includes a set of measures, both objective and subjective, which can be applied and adapted to a variety of situations. The objective measures derive from the original Buckley et al., (1983) work. The subjective measures include 1) observer evaluations using the behavioral rating scales (Sollenberger et al., 1997; Vardaman & Stein, 1998) and 2) controller ratings by simulation participants. The controller ratings include measures of workload, situational awareness, and project specific issues.

1.1.5 Validity and Reliability of Tower Cab Measures

Any new measures proposed for the Tower Cab environment will need to have both validity and reliability to be useful for researchers. A measurement has validity if it measures the characteristic that it is supposed to measure. Reliability refers to the consistency of the measure. When measuring job performance in applied settings, it is critical to conduct appropriate assessments to establish the validity and reliability of the measures.

To achieve measurement validity, researchers must first gain an understanding of air traffic control and the job controllers perform. Researchers can accomplish this by conducting a job task analysis, by observing controllers at actual facilities or in simulation, discussing the job with controllers, or reading texts describing their duties. When developing new measures, it is important to actively involve ATCSs in the process and draw upon their knowledge and expertise. Experienced ATCSs are experts in observing, identifying, and evaluating controller performance. The more ATCSs that can be involved in the development process the better it will be for the validity of a new measure. However, one problem for establishing measurement validity through the opinions of ATCSs is that experts do not always agree on what are the important aspects of performance or what levels indicate satisfactory performance.

There are several different types of validity. Face validity and content validity are similar constructs. Face validity is simple, subjective, and not a very scientific standard for achieving "true" validity. Face validity is achieved when the developers of a measure agree that their proposed measure is relevant to job performance. Content validity is stricter and is achieved when the contents are systematically assessed, usually with experts in the job (e.g., ATCSs, supervisors, and training experts) who agree that the proposed measurement tool and all its components are relevant to job performance.

Content validity has been a minimum standard for development of previous ATC measures (Sollenberger et al., 1997). One method used at the RDHFL for ensuring content validity is for researchers to team with several subject matter experts (SMEs) in a workshop to discuss and determine the relevant aspects of job performance. The research team can then develop measures (e.g., rating scales) that evaluate these components of job performance. To have content validity, experts should agree that every component of the proposed measurement tool is relevant to job performance and that all relevant aspects of job performance are represented in the tool. If there is substantial disagreement among the experts, then the individual component may have to be refined or omitted from the measurement tool. Any relevant aspect of job performance that is not represented should be added to the measurement tool to achieve content validity.

Predictive, concurrent, and construct validity are types of validity that researchers can evaluate by scientific methods using statistical correlation, such as the Pearson method. Predictive and concurrent validity both refer to a measure's ability to predict or agree with a specific criterion measure. The proposed tool does not have to directly measure the same construct as the criterion measure. Examples are: scores on a test of spatial ability may predict successful ATC ability, or measures of arithmetic skills may agree with successful completion of training problems. Predictive validity can be determined by correlating measures from a proposed tool with criterion measures obtained while evaluating later performance. Concurrent validity can be determined by correlating measures from a proposed tool with criterion measures taken at the same time while evaluating performance. Construct validity is usually considered the strictest form of validity and refers to a measure's ability to correlate with other closely related constructs or an accepted standard measurement of performance. For example, a proposed rating form for assessing controller performance in simulation research would have good construct validity if it has a high correlation with a standard evaluation form used at field facilities during on-the-job training. Ideally, new metrics would be assessed against the criteria.

Any proposed performance measure for the Tower Cab environment must have good reliability as well as validity to be useful to researchers in a simulation environment. Two types of reliability are important for a proposed measurement tool based upon subjective ratings: intra-rater and inter-rater. Intra-rater reliability refers to the measure's consistency when used by the same evaluator to measure performance at different times. Inter-rater reliability refers to the measure's consistency when used by different evaluators to measure the same level of performance.

Researchers can use statistical correlation to determine both intra-rater and inter-rater reliability. For intra-rater reliability, measures must be obtained from the same evaluators at different times, and the strength of the correlation between the ratings determines the degree of reliability. Intra-rater reliability is based upon within-subject variability of ratings and can be computed by the Pearson correlation method. For inter-rater reliability, measures must be obtained from different evaluators, preferably at the same time. Inter-rater reliability is based upon between-subject variability of ratings and can be computed by either the Pearson or intraclass correlation methods. We have assessed both types of reliability for the previous rating scales developed at the RDHFL for the en route and TRACON environments (Sollenberger et al., 1997; Vardaman & Stein, 1998).

2. Candidate Measures for the Tower Cab Environment

To identify candidate measures, we held preliminary discussions with SMEs and project sponsors and reviewed existing literature. As in most environments, some measures of performance exist and some need to be developed. Within the Tower Cab environment, for example, there are some existing measures for evaluating controller performance, such as field over-the-shoulder evaluations. However, we needed to develop behavioral ratings for the Tower Cab that could be used for either local or ground control positions. Our goal was to develop a set of measures that fit the RDHFL human factors model of objective and subjective ratings. The following sections review the existing objective and subjective measures and our development of behavioral rating scales.

2.1 Objective Measures

Objective measures exist in the Tower Cab environment and can generally be classified as safety, capacity, or efficiency measures. A few safety measures are already in common usage in Towers, such as operational errors/deviations (e.g., runway incursions). Buckley et al., (1983) and Hadley et al., (1999) identified a number of others, and we will discuss these in our selected measures section.

There are also capacity measures that are in common usage, such as the number of aircraft takeoffs and landings per hour. These numbers have been used to establish staffing requirements and, when adjusted for weather conditions and traffic mix, may give an indication of controller performance or equipment effectiveness. Such measures could be used as indicators of system efficiency within the research context.

Other efficiency measures, such as the number and duration of communications, are less obvious performance measures because they usually do not directly impact passenger safety or airport capacity. However, communications do affect controller workload. When controller workload reaches high enough levels, aircraft delays may occur or safety may become compromised. Controller communications are an important source of information to evaluate controller performance and system effectiveness in the Tower Cab environment. For example, effective local/ground controller teams are able to anticipate each other's actions and minimize coordination communications. Also, good anticipation between team members may allow a ground controller to quickly taxi an aircraft across a runway instead of requiring the aircraft to hold short and wait.

Another candidate measure for the Tower Cab environment is the "hourly classification index," which is currently used to classify terminal facilities and determine the pay scale for ATCSs (FAA, 1998b). The hourly classification index measures the overall activity for a specific facility as well as the general taskload demands on controllers at that facility. The formula for the hourly classification index is based upon the average number of flight operations at a facility and a complexity factor known as the "sustained traffic index." Factors such as runway configurations, traffic mix, number of IFR/VFR arrivals, departures, and military flights are given specified weights and combined in the formula. Although the hourly classification index is a good measure of the taskload demands on controllers at individual facilities, it is not a human factors measure of performance that can be used in simulation research.

2.1.1 Selection of Measures for the Tower Cab Environment

After reviewing publications and review sources of previous aviation human factors research, we selected a set of objective measures for use in Tower Cab simulations. These measures include SEMs described by Buckley et al., (1983) and Stein and Buckley (1992) that measure ATC safety, capacity, and efficiency. We also included measures and some techniques for performance assessment from the Hadley et al., (1999) database (e.g., communication frequency and duration). We did not intend the selected set of measures to be a complete compilation of measures from these previous sources. However, we selected the measures because they are applicable and important to the Tower Cab environment and are practical to use in simulation research. These measures meet the goal of continuity with measures that have been used in previous TRACON and en route simulations and represent a range of overall system effectiveness and human factors performance measures. Having been used in previous research, the measures have some validity and reliability and should be useful in the Tower Cab environment. The proposed objective measures are presented in Table 1.

2.2 Subjective Ratings

Subjective ratings are a key component of the RDHFL's ATC human factors measurement model. Two sources of subjective ratings are from an expert observer and the participating controller. Some existing measures for the controller ratings, such as the NASA Task Load Index, can be transitioned to Tower Cab; however, we must develop or modify additional measures. For example, the Behavioral Rating Form required modification for the Tower Cab environment. The following sections describe the development of the observer rating scale and existing measures that can be adapted.

Table 1. Objective Performance Measures for the Tower Cab Environment

I – Safety Measures
RI – Frequency of runway incursions
RID – Duration of runway incursions
LCNF – Frequency of longitudinal conflicts for aircraft on approach
LCNFD – Duration of longitudinal conflicts for aircraft on approach
PCNF – Frequency of conflicts for aircraft on simultaneous parallel approach
PCNFD – Duration of conflicts for aircraft on simultaneous parallel approach
ASCNF – Frequency of intrusions into restricted airspace
ASCNFD – Duration of intrusions into restricted airspace
CA – Frequency of conflict alerts
CPA – Closest point of approach for each conflict
CPAHSEP – Horizontal separation at CPA time
CPAVSEP – Vertical separation at CPA time

II – Capacity Measures
NFLT – Number of flights handled during simulation run
LAND – Number of landings during simulation run
DEPART – Number of departures during simulation run
HOIN – Number of handoffs received during simulation run
HOUT – Number of handoffs given during simulation run
NMAX – Maximum number of aircraft controlled at same time during simulation run

III – Efficiency Measures
NPTT – Frequency of controller push-to-talk communications
DPTT – Duration of controller push-to-talk communications (cumulative)
TIME – Average time an aircraft spent under controller’s control
DIST – Average distance flown by an aircraft during simulation
FUEL – Average fuel used by an aircraft during simulation
AVLAND – Average time interval between landing aircraft
AVDEPART – Average time interval between departing aircraft
NMISS – Number of missed approaches

2.2.1 Development of a Behavioral Rating Form for the Tower Cab Environment

We identified a behavioral rating form as one of the products for the Tower Cab Metrics project. There are, however, some rating scales in use at field facilities that could guide the development of rating scales for research settings in addition to the CTA Task Analysis (Alexander et al., 1989). One such form is used to rate developmental controllers as they acquire skills in the occupation. The "ATCT/ARTCC OJT Instruction/Evaluation Report" (FAA, 1998a) evaluates job task performance within six categories: 1) separation, 2) coordination, 3) control judgment, 4) methods and procedures, 5) equipment, and 6) communication. This form provides for ratings of satisfactory, needs improvement, or unsatisfactory (see Appendix A).

For this project, we developed a behaviorally based rating form for the Tower Cab environment that was similar to the rating forms already developed at the FAA WJHTC for the TRACON and en route environments (Sollenberger et al., 1997; Vardaman & Stein, 1998). Following the same development process that was completed for the previous rating forms, we identified the performance categories and observable behaviors for local and ground controllers. We accomplished this through observing operations at major ATCTs, obtaining on-the-job training reports and checklists from field facilities, obtaining position descriptions and responsibilities from ATC manuals (FAA, 1997), and interviewing SMEs in the Tower Cab environment. The proposed behavioral rating form has 25 individual rating scales for the Tower Cab environment (see Appendix B).

We designed the rating form as a comprehensive tool to assess the critical tasks of local and ground controllers. The assistance of experienced ATCSs was important in determining these critical tasks and provides the rating form with face validity as a tool to measure controller performance. The rating form also identifies observable controller behaviors for each rating scale that should reduce potential confusion amongst users and improve rating accuracy. We adopted an 8-point rating systems that should allow enough sensitivity to evaluate new ATC systems, procedures, or advanced concepts in a simulation environment (Sollenberger et al., 1997). As in the previous rating forms, we provided an observations section for each rating scale for ATCSs to comment about specific behaviors identified.

2.2.1.1 Training for ATCSs using the Behavioral Rating Form

User training is an important process that must be completed to ensure ratings are consistent and reliable. First, users should be ATCSs who are experienced in observing and evaluating controller performance. Usually supervisors are the best choice, however, senior ATCSs who have experience training other controllers are also good candidates. Inexperienced users may find it difficult to identify the appropriate behaviors and evaluate the proficiency of the controllers conducting traffic. However, even experienced ATCSs must become familiar with the rating form by using it several times in "hands on" sessions while observing controllers in action. In a simulation setting, this may be accomplished by using the rating form during several practice scenarios before rating controllers during actual test scenarios. The more training and practice ATCSs can obtain with the rating form, the better the results should be for achieving reliability and making conclusions about the ideas being tested in simulation.

When two or more controllers operate different sectors or simultaneously participate in parallel sessions of the simulation, multiple ATCSs are often required to make observations and ratings.

In order to ensure consistency and reliability amongst ATCSs using the rating form, the users must discuss and agree on the standards that they will be using to evaluate controller performance. Some ATCSs may have high standards of performance, whereas others may not be as strict. It is important that ATCSs adopt mutual performance criteria while using the rating form. Again, this will usually require using the rating form during several practice sessions where all users observe the same controller conducting traffic. The users should be encouraged to discuss amongst themselves the controller behaviors they observed and how they would rate the behaviors using the scales. If practical, multiple ATCSs should observe each controller and assess the inter-rater reliability of ratings.

Future research should assess the reliability of the proposed rating form using a study similar to previous research evaluating the rating forms for the TRACON and en route environments (Sollenberger et al., 1997; Vardaman & Stein, 1998). Researchers should select a group of ATCSs that will view videotapes of controllers conducting traffic during Tower Cab simulations. After the appropriate training and practice, the ATCSs will use the rating form to evaluate controller performance in the videotapes. The ATCSs will view a few of the same videotapes twice, on separate days to evaluate the intra-rater reliability of the rating form. We will also assess the inter-rater reliability of the rating form by comparing the ATCSs ratings of the same videotape. For the purpose of reliability testing, watching videotapes is preferable to observing "live" tower operations because videotapes can be presented twice, and all ATCSs will view the same controller behaviors.

2.2.2 Existing Subjective Measures for Controller Ratings

After reviewing existing measures, we selected the following workload and situational awareness measures as candidates for the core set of Tower Cab Metrics.

Table 2. Performance Measures for the Tower Cab Environment

IV – Workload and Situation Awareness Measurement Techniques
ATWIT – Air Traffic Workload Input Technique: A uni-dimensional workload technique that obtains instantaneous workload ratings as controllers work traffic (Stein, 1985). NASA TLX – NASA Task Load Index: A multi-dimensional workload technique that obtains post-scenario workload ratings from controllers (Hart & Staveland, 1987). SWAT – Subjective Workload Assessment Technique: A multi-dimensional workload technique that obtains post-scenario workload ratings from controllers (Redi & Nygren, 1988). SAGAT – Situational Awareness Global Assessment Technique: An objective situational awareness technique that involves "freezing" scenarios and asking controllers questions about the air traffic situation at the time of the freeze (Endsley, 1995). SART – Situation Awareness Rating Technique: A subjective situational awareness technique that involves post-scenario ratings from controllers along several scales (Selcon & Taylor, 1989).

3. Discussion

We only intend this initial effort to develop candidate performance measures for the Tower Cab environment. The result is a battery of human factors metrics that include objective measures (e.g., number of errors, number of aircraft handled) and subjective measures for observers (e.g., behavioral ratings) and participating controllers (e.g., workload and situational awareness). This proposed set of measures is representative of the core set used by RDHFL researchers in the other environments. In addition, the measures are applicable to both ground and local control positions. As a whole, these measures provide a broad look at performance for researchers to apply in a simulation environment. This allows for conducting experiments to obtain both baseline and comparison measures for assessing the effects of any proposed changes in technology, automation, or procedures.

In addition to the objective and subjective measures proposed, we recommend that researchers include other measures that are specific to the ideas being evaluated in the simulation. For example, if the objective of the simulation is to evaluate a new communications system, then we should include measures that assess the signal quality, intelligibility, and acceptability to users. These more specific measures are beyond the scope of the present report because more would need to be known about the specific objectives of the individual tests and evaluations. Finally, measures of visual scanning or psychophysiology may be considered according to the requirements of the study.

Also, we must conduct more research to assess the validity and reliability of the proposed measures in the Tower Cab environment through continued usage in simulations. We have described a specific method from previous research that can be used to assess the reliability of the proposed behavioral rating form. However, we need further research to evaluate the proposed objective measures for the Tower Cab environment.

This review of existing measures revealed very few surface movement metrics. This would leave adequate measures of ground controller’s performance incompletely addressed. It is possible that, in many towers, the ground controller has the more complex and difficult job. Future research should examine these issues as well as develop metrics for the remaining positions in the Tower Cab.

References

Alexander, J. R., Alley, V. L., Ammerman, H. L., Fairhurst, W. S., Hostetler, C. M., Jones, G. W., & Rainey, C. L. (1989). *FAA Air Traffic Control operations concepts Volume VII: ATCT tower controllers* (DOT/FAA/AP87/01). Atlantic City International Airport, NJ: DOT/FAA Technical Center.

Buckley, E. P., DeBaryshe, B. D., Hitchner, N., & Kohn, P. (1983). *Methods and measurements in real-time air traffic control system simulation* (DOT/FAA/CT-83/26). Atlantic City International Airport, NJ: DOT/FAA Technical Center.

Endsley, M. R. (1995). Measurement of situation awareness in dynamic systems. *Human Factors*, *37*(1), 65-84.

FAA. (1997). *Philadelphia tower standard operating procedures* (FAA Order 7110.23B). Washington, DC: Federal Aviation Administration.

FAA. (1998a). *ATCT/ARTCC OJT Instruction/Evaluation Report* (FAA Form 3120-25). Washington, DC: Federal Aviation Administration.

FAA. (1998b). *Position classification standard for Air Traffic Control series ATC-2152 terminal and en route*. Washington, DC: Federal Aviation Administration.

Hadley, G. A., Guttman, J. A., & Stringer, P. G. (1999). *Air Traffic Control Specialist performance measurement database* (DOT/FAA/CT-TN99/17). Atlantic City International Airport, NJ: DOT/FAA William J. Hughes Technical Center.

Hart, S. G., & Staveland, L. E. (1987). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In P.A. Hancock and N. Meshkati (Eds.), *Human Mental Workload*. Amsterdam: North-Holland.

Redi, G. G., & Nygren, T. E. (1988). The subjective workload assessment technique: A scaling procedure for measuring mental workload. In P. A. Hancock & N. Meshtaki (Eds.), *Human Mental Workload*. Amsterdam: North-Holland.

Selcon, S. J., & Taylor, R. M. (1989). *Evaluation of the situational awareness rating technique (SART) as a tool for aircrew systems design (*AGARD-CP-478). Neuilly Sur Seine, France: NATO-AGARD.

Sollenberger, R. L., Stein, E. S., & Gromelski, S. (1997). *The development and evaluation of a behaviorally based rating form for the assessment of air traffic controller performance* (DOT/FAA/CT-TN96/16). Atlantic City International Airport, NJ: DOT/FAA William J. Hughes Technical Center.

Stein, E. S. (1985). *Air traffic controller workload: An examination of workload probe* (DOT/FAA/CT-TN84/24). Atlantic City International Airport, NJ: DOT/FAA Technical Center.

Stein, E. S., & Buckley, E. P. (1992). *Simulation Variables*. Atlantic City International Airport, NJ: DOT/FAA Technical Center, NAS Human Factors Branch (ACT-530). Unpublished document.

Stein, E. S., & Buckley, E. P. (1994). *Human Factors at the FAA Technical Center: Bibliography 1958-1994* (DOT/FAA/CT-TN94/50). Atlantic City International Airport, NJ: DOT/FAA Technical Center.

Vardaman, J. J., & Stein, E. S. (1998). *The development and evaluation of a behaviorally based rating form for the assessment of en route air traffic controller performance* (DOT/FAA/CT-TN98/5). Atlantic City International Airport, NJ: DOT/FAA William J. Hughes Technical Center.

Wickens, C. D., Mavor, A. S., & McGee, J. P. (1997). *Flight to the future, human factors in air traffic control.* Washington, D.C.: National Academy Press.

Appendix A

ATCT/ARTCC OJT Instruction/Evaluation Report – Cover Page

1. Name		2. Date	3. Scenario/Position(s)
4. Weather ☐ VFR ☐ MVFR ☐ IFR ☐ Other	5. Workload ☐ Light ☐ Moderate ☐ Heavy	6. Complexity ☐ Not Difficult ☐ Occasionally Difficult ☐ Mostly Difficult ☐ Very Difficult	7. Hours
			8. Total Hours This Position
9. Purpose ☐ OJT ☐ OJF ☐ Familiarization Scenario ☐ Instructional Scenario ☐ Evaluation Scenario ☐ Skill Check ☐ Certification ☐ Recertification ☐ Skill Enhancement ☐ Other			10. Routing

11.	Job Task	Job Subtask	Observed	Comment	Satisfactory	Needs Improvement	Unsatisfactory	Simulation Training
Performance	A. Separation	1. Separation is ensured.						
		2. Safety alerts are provided.						
	B. Coordination	3. Performs handoffs/pointouts.						
		4. Required coordinations are performed.						
	C. Control Judgment	5. Good control judgment is applied.						
		6. Priority of duties is understood.						
		7. Positive control is provided.						
		8. Effective traffic flow is maintained.						
	D. Methods and Procedures	9. Aircraft identity is maintained.						
		10. Strip posting is complete/correct.						
		11. Clearance delivery is complete/correct and timely.						
		12. LOAs/directives are adhered to.						
		13. Additional services are provided.						
		14. Rapidly recovers from equipment failures and emergencies.						
		15. Scans entire control environment.						
		16. Effective working speed is maintained.						
	E. Equipment	17. Equipment status information is maintained.						
		18. Equipment capabilities are utilized/understood.						
	F. Communication	19. Functions effectively as a radar/tower team member.						
		20. Communication is clear and concise.						
		21. Uses prescribed phraseology.						
		22. Makes only necessary transmissions.						
		23. Uses appropriate communications method.						
		24. Relief briefings are complete and accurate.						
	G. Other							

Appendix B

Subject Matter Expert Observer Rating Form

Tower Cab Version

Observer Code _________ Date _________

Controller _________ Scenario _________

INSTRUCTIONS

This form is designed to be used by supervisory air traffic control specialists to evaluate the effectiveness of controllers working in simulation environments. SATCSs will observe and rate the performance of controllers in several different performance dimensions using the scale below as a general purpose guide. Use the entire scale range as much as possible. You will see a wide range of controller performance. Take extensive notes on what you see. Do not depend on your memory. Write down your observations. Space is provided after each scale for comments. You may make preliminary ratings during the course of the scenario. However, wait until the scenario is finished before making your final ratings and remain flexible until the end when you have had an opportunity to see all the available behavior. At all times please focus on what you actually see and hear. This includes what the controller does and what you might reasonably infer from the actions of the pilots. Try to avoid inferring what you think may be happening. If you do not observe relevant behavior or the results of that behavior, then you may leave a specific rating blank. Also, please write down any comments that may help improve this evaluation form. Do not write your name on the form itself. Your identity will remain anonymous, as your data will be identified by an observer code known only to yourself and the researchers conducting this study. The observations you make do not need to be restricted to the performance areas covered in this form and may include other areas that you think are important.

ASSUMPTIONS

ATC is a complex activity that contains both observable and unobservable behavior. There are so many complex behaviors involved that no observational rating form can cover everything. A sample of the behaviors is the best that can be achieved, and a good form focuses on those behaviors that controllers themselves have identified as the most relevant in terms of their overall performance. Most controller performance is at or above the minimum standards regarding safety and efficiency. The goal of the rating system is to differentiate performance above this minimum. The lowest rating should be assigned for meeting minimum standards and also for anything below the minimum since this should be a rare event. It is important for the observer/rater to feel comfortable using the entire scale and to understand that all ratings should be based on behavior that is actually observed.

Rating Scale Descriptors

Remove this Page and keep it available while doing ratings

SCALE	QUALITY	SUPPLEMENTARY
1	**Least Effective**	Unconfident, Indecisive, Inefficient, Disorganized, Behind the power curve, Rough, Leaves some tasks incomplete, Makes mistakes
2	**Poor**	May issue conflicting instructions, Doesn't plan completely
3	**Fair**	Distracted between tasks
4	**Low Satisfactory**	Postpones routine actions
5	**High Satisfactory**	Knows the job fairly well
6	**Good**	Works steadily, Solves most problems
7	**Very Good**	Knows the job thoroughly, Plans well
8	**Most Effective**	Confident, Decisive, Efficient, Organized, Ahead of the power curve, Smooth, Completes all necessary tasks, Makes no mistakes

I - MAINTAINING SAFE AND EFFICIENT TRAFFIC FLOW

1. Maintaining Separation and Resolving Potential Conflicts........... **1 2 3 4 5 6 7 8**
   - using control instructions that maintain appropriate aircraft and runway separation
   - detecting and resolving impending conflicts early
   - using proper wake turbulence separation

   Observations:

2. Sequencing Aircraft Efficiently... **1 2 3 4 5 6 7 8**
   - using efficient and orderly spacing techniques for arrival and departure aircraft
   - maintaining safe arrival and departure intervals that minimize delays

   Observations:

3. Using Control Instructions Effectively/Efficiently........................ **1 2 3 4 5 6 7 8**
   - providing accurate taxi/hold short instructions to pilots
   - issuing economical clearances that result in need for few additional instructions to handle aircraft completely

   Observations:

4. Overall Safe and Efficient Traffic Flow Scale Rating................... **1 2 3 4 5 6 7 8**

   Observations:

II - MAINTAINING ATTENTION AND SITUATION AWARENESS

5. Maintaining Awareness of Aircraft Positions................................ **1 2 3 4 5 6 7 8**
   - scanning runways and taxiways for aircraft and ground vehicles prior to issuing take-off and landing clearances

   Observations:

6. Ensuring Positive Control.. **1 2 3 4 5 6 7 8**
   - tailoring control actions to situation
   - using effective procedures for handling heavy, emergency, and unusual traffic and airport situations

   Observations:

7. Detecting Pilot Deviations from Control Instructions.................... **1 2 3 4 5 6 7 8**
   - ensuring that pilots follow assigned clearances correctly
   - correcting pilot deviations in a timely manner

   Observations:

8. Correcting Own Errors in a Timely Manner.................................. **1 2 3 4 5 6 7 8**
   - acting quickly to correct errors
   - changing an issued clearance when necessary to expedite traffic flow

   Observations:

9. Overall Attention and Situation Awareness Scale Rating **1 2 3 4 5 6 7 8**

   Observations:

III – PRIORITIZING

10. Taking Actions in an Appropriate Order of Importance................**1 2 3 4 5 6 7 8**
    - resolving situations that need immediate attention before handling lower priority tasks
    - issuing control instructions in a prioritized, structured, and timely manner

    Observations:

11. Preplanning Control Actions.. **1 2 3 4 5 6 7 8**
    - recognizing conflicting traffic in sufficient time to take control actions and avoid last second decisions
    - studying overall traffic flow on the airport to adjust operations as necessary to avoid confusion and delays

    Observations:

12. Handling Control Tasks for Several Aircraft................................. **1 2 3 4 5 6 7 8**
    - shifting control tasks between several aircraft when necessary
    - communicating in timely fashion while sharing time with other actions

    Observations:

13. Marking Flight Strips while Performing Other Tasks **1 2 3 4 5 6 7 8**
    - marking flight strips accurately while talking or performing other tasks
    - keeping flight strips current

    Observations:

14. Overall Prioritizing Scale Rating.. **1 2 3 4 5 6 7 8**

    Observations:

IV – PROVIDING CONTROL INFORMATION

15. Providing Essential Air Traffic Control Information...................... **1 2 3 4 5 6 7 8**
    - providing mandatory services and advisories to pilots in a timely manner
    - exchanging essential information

    Observations:

16. Providing Additional Air Traffic Control Information.................. **1 2 3 4 5 6 7 8**
    - providing additional services when workload is not a factor
    - exchanging additional information

    Observations:

17. Providing Coordination.. **1 2 3 4 5 6 7 8**
    - providing effective and timely coordination
    - using proper point-out procedures

    Observations:

18. Overall Providing Control Information Scale Rating....................**1 2 3 4 5 6 7 8**

    Observations:

V – TECHNICAL KNOWLEDGE

19. Showing Knowledge of LOAs and SOPs....................................... **1 2 3 4 5 6 7 8**
    - controlling traffic as depicted in current LOAs and SOPs
    - performing transfer of aircraft control procedures correctly

    Observations:

20. Showing Knowledge of Aircraft Capabilities and Limitations....... **1 2 3 4 5 6 7 8**
    - using appropriate control instructions to separate aircraft with varied flight capabilities and wake vortex restrictions
    - issuing clearances that are within aircraft performance parameters

    Observations:

21. Overall Technical Knowledge Scale Rating................................... **1 2 3 4 5 6 7 8**

    Observations:

VI – COMMUNICATING

22. Using Proper Phraseology.. **1 2 3 4 5 6 7 8**
    - using words and phrases specified in the 7110.65
    - using phraseology that is appropriate for the situation
    - using minimum necessary verbiage

    Observations:

23. Communicating Clearly and Efficiently **1 2 3 4 5 6 7 8**
    - speaking at the proper volume and rate for pilots to understand
    - speaking fluently while scanning or performing other tasks
    - ensuring clearance delivery is complete, correct and timely
    - speaking with confident, authoritative tone of voice

    Observations:

24. Listening to Pilot Readbacks and Requests **1 2 3 4 5 6 7 8**
    - correcting pilot readback errors
    - acknowledging pilot or other controller requests promptly
    - processing requests correctly in a timely manner

    Observations:

25. Overall Communicating Scale Rating **1 2 3 4 5 6 7 8**

    Observations: