

# Cognitive Workload and Visual Attention Analyses of the Air Traffic Control Tower Flight Data Manager (TFDM) Prototype Demonstration

Kiran Lokhande & Hayley J. Davison Reynolds  
MIT Lincoln Laboratory

This paper presents two methods of analyzing air traffic controller activity: cognitive workload measurement through the novel comparison of controller-pilot verbal communications, and visual attention quantification through manual eye gaze analysis. These analyses were performed as part of an evaluation of the Tower Flight Data Manager (TFDM) prototype system. Cognitive workload analyses revealed that, when comparing participant controllers utilizing TFDM to a control group utilizing existing air traffic control (ATC) equipment, participants issued commands sooner than the control, and thus were perceived to have a lower workload. While visual attention data were not available for the control group, analyses of participant gaze data revealed 81.9% of time was spent in a head-down position, and 17.2% of time was spent head-up. Results are related back to system inefficiencies to find potential areas of improvement in design.

## INTRODUCTION

Air traffic controllers face the demanding task of managing daily flight arrivals and departures at a given airport. On top of this demand, they currently deal with systems that are slowly approaching their capacity and limit of improvement (FAA, 2012). One source of increased cognitive workload is the lack of integrated information sources and the difficulty this presents when accompanied by rising flight demand (FAA, 2012). While ongoing development of technology aims to consolidate these information systems and improve efficiency in the tower environment, research has also focused on finding proper ways to measure behavioral indicators, such as cognitive workload and visual attention, to inform the design of new, integrated system designs.

### Cognitive workload measurement in ATC

Cognitive workload measurement methods within the Air Traffic Control (ATC) domain can be heavily subjective, intrusive, and insensitive to individual differences. Many ATC studies have utilized subjective or self-report measures of workload (e.g., Metzger & Parasuraman, 2001; Endsley & Rodgers, 1997). While potentially useful, self-report ratings are susceptible to participant bias (Neale & Liebert, 1973; Howard & Dailey, 1979). On the other hand, more objective measures of workload have their own complications. Psychophysiological measures recorded through electroencephalography (EEG) equipment, for example, are highly accurate, but EEG data analysis requires researcher expertise (Wilson & Russell, 2003; Berka et al., 2007). EEG headgear, often requiring electrodes applied directly to participants' scalps, is also rather intrusive. In 2004, Averty et al. began to bridge the gap between subjective and objective measures of controller workload. Their study quantified "emotional factors" such as *seriousness* and *urgency* based on aircraft distance and relative flight paths between conflicting aircraft to create an overall traffic load index (TLI) (Averty et al., 2004). However, these measures did not consider physical or quantifiable aspects of the controllers themselves. The aforementioned studies did not assess individual differences when evaluating workload and/or had no way of measuring results against a control. There was also no clear method of pinpointing specific workload issues or errors committed by controllers.

Comparatively, controller-pilot communications provide an objective measure that is inclusive of controller differences. Factors such as content/error-incidence, and length and number of communications correlate with workload (e.g., Embrey, Blackett, Marsden, & Peachey, 2006; Cardosi, 1993). Data collection of communications, through observation or audio recordings, is also cost-effective and minimally intrusive.

Measuring controller-pilot communications is a performance-based measure, used as a primary task in earlier studies (Wierwille & Eggemeier, 1993). Another method of measuring mental capacity and load with performance-based measures is through secondary tasks (Wickens, 1992; Funke, Knott, Salas, Pavlas, & Strang, 2012). Secondary tasks are useful for assessing workload changes through measuring spare mental capacity. They are also useful when the task is one naturally performed by the participant (Wierwille & Eggemeier, 1993). For example, Metzger & Parasuraman (2001) tested controller ability to detect aircraft conflicts under varying traffic load. To measure mental workload, an embedded secondary task of monitoring flight progress and updating flight strips was given.

### Visual attention measurement in ATC

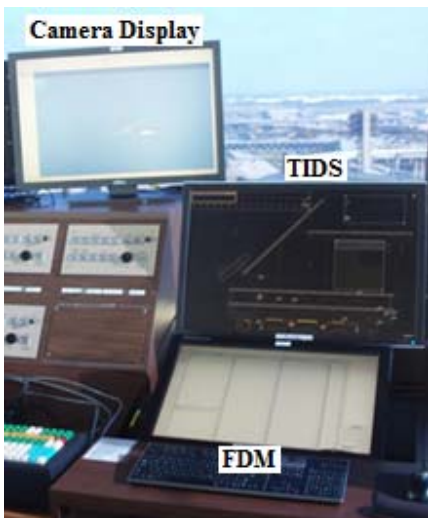
Eyetracking is another objective method of research that reveals behavioral acclimation to technology. Several earlier ATC studies have noted the importance of attentional division, since an overreliance on technology leaves a controller susceptible to decreased situation awareness and thus potential error (Parasuraman & Riley, 1997; Endsley & Rodgers, 1997). In addition to preventing cognitive overload, another crucial goal of an ATC system is to enable a controller to divide attention between "head-down" displays inside the tower and "head-up" information seen outside of the tower to maintain an accurate awareness of current conditions (Endsley & Rodgers, 1997). Such a balance is important since both head-down and head-up information provide unique ATC insights. Controllers are procedurally required to separate aircraft using out-the-window information, which makes the ability to remain head-up of utmost importance to the safety of tower operations. In 1996, Bruce cited an average of 35-38% head-down time when observing controller visual attention in several US towers (as cited in Cardosi & Yost, 2001). Other European studies have reported varying head-down measures: 80%

(Pavet, 2001), 35-39% (Pinska, 2007), and 58-60% (SensoMotoric Instruments, 2011). While each study's tower environment is different, results demonstrate a common trend of reliance upon technical displays over head-up information.

## METHOD

In this study, controller-pilot communications are used to measure cognitive workload, and visual attention is measured through *manual gaze analysis* to eliminate the need for cumbersome, expensive equipment. Through analyzing both cognitive workload and visual attention, this exemplifies the role human behavioral measurement can have in improving the design of the Tower Flight Data Manager prototype system (TFDM). TFDM was designed to aid the management of surface operations and to improve information exchange and enhance operational efficiency. Simultaneously, TFDM aimed to avoid unnecessary increase in workload or extensive requirement of visual attention. These motives prompted the use of novel performance assessment measures to quantify workload and attention when using the TFDM prototype.

At the time of testing, TFDM had undergone one previous field demonstration at Dallas / Ft. Worth International Airport (DFW). For this second demonstration, also at DFW, three main components of the system were tested, each one supplementing a specific part of ATC processes: Flight Data Manager (FDM), Tower Information Display System (TIDS), and COHU™ camera, each seen in Figure 1. Participants were given access to these displays in addition to their standard,



**Figure 1: TFDM components.**

everyday equipment. The FDM automated the traditional paper flight strip system and enabled the sequencing and tracking of flights on a touch-screen display. This display had unique configurations for ground and local controllers, each according to their specific needs and duties. The TIDS provided enhanced ground and air surveillance

processing capabilities. Lastly, the COHU™ was a video camera used to automatically track flights. It could also be viewed on the TIDS in an embedded window. Participants were twelve controllers with an average of 17.6 years (min: 0.5 years, max: 32 years) of controller experience. In performance assessments, two participants a day were observed using the TIDS, FDM, and COHU™ displays, from appx. 7am to 2:45pm. They were first trained in use of TFDM, and then each one spent half of their day as a ground controller and the other half as a local controller. Participants were instructed to evaluate and simulate control of air traffic,

while utilizing and simultaneously evaluating the functionality of the TFDM prototype. It should be noted that the majority of these participants had been involved in the first field demonstration, and thus already had a basic knowledge of the system.

Participants were tested in the Center Tower of DFW and issued verbal instructions to pilots as if they were in control of the aircraft. Meanwhile, controllers in the East Tower served as the control group and issued actual instructions and retained aircraft separation responsibility. Participants could hear the instructions of East Tower controllers (ETCs), though ETCs could not hear them. Participant audio/video, ETC/pilot audio, and live-feed screen data were recorded and compiled onto one synchronized playback system for performance analyses (Figure 2). Performance measures were then subsequently correlated back to the original video/audio to find design flaws



**Figure 2: Video/audio playback system used for post-demonstration analyses.**

and avenues of TFDM display improvement.

### TFDM Cognitive workload measurement

During the first DFW field demonstration, researchers noted that participants often issued instructions to pilots before their ETC counterparts (Davison Reynolds, Kuffner, & Yenson, 2011). When issuing instructions later, participants often appeared to be confused about a TFDM feature or distracted by a workload-inducing situation. These notes motivated the consideration that comparing instruction times between ETCs and participants might serve as a method of measuring participant cognitive workload. While other studies clearly indicate several relationships between controller-pilot communications and workload (e.g., Embrey et al., 2006; Cardosi, 1993), none have yet compared instruction issuance between controllers. These studies support the notion that a longer participant response time could indicate a higher cognitive workload. When further analyzing instruction issuances via video, these measures also heavily aided in identifying TFDM design issues.

In the present study, comparing instruction issuance presented an objective method of measuring cognitive workload. Controller-pilot communications were quantifiably compared *between* participants using the TFDM and ETC control subjects operating in a typical ATC setup. Similar to Metzger &

Parasuraman (2001), this was treated as a secondary task. Since participants were instructed to primarily use and evaluate the TFDM system, instruction issuance was considered a secondary task. Verbal instruction issuance times were compared between these controllers, thus avoiding subjectivity. Since microphones and recordings of controller-pilot transmissions were the only data necessary, data collection was also cost-effective and minimally intrusive.

Through video and audio analyses, participant and ETC verbal instructions were compared and two quantitative measures were used to determine cognitive workload: *gap time* between participant and ETC instructions, and *response rate*, i.e. the percentage of time participants issued an instruction before, after, or within one second of/neutral to their ETC counterpart. An onscreen timer placed in the center of each data-analysis video helped record the time of each verbal instruction (Figure 2). While the onscreen timer was precise up to hundredths of a second, events were recorded at the level of 1 s, due to the potentially imprecise nature of data collection.

### TFDM Visual attention measurement

In DFW video data, indications of participant eye direction, head tilt, onscreen mouse movement, and verbal contextual information enabled researchers to determine the focus of participant attention at any given point. To analyze participant visual attention, video data were sampled to manually record the location and duration of these visual gazes. The duration of each gaze was noted through use of the on-screen timer (to the precision of 1 s). For consistency, manual gaze analysis was performed by one researcher and their accuracy was calibrated separately before analyses. After gathering data, numerical analyses were performed on *percentage of total dwell time* spent on individual displays and head-up areas.

## RESULTS

### Cognitive workload results

To compare data between participants and the ETC control group, verbal command data were analyzed through two-way Welch t-tests due to uneven sample sizes and potentially unequal population variances. Data were also analyzed within subjects with paired t-tests to examine differences between participants' respective ground and local controller roles.

The results indicate that participants often issued instructions sooner than ETCs, and local participant controllers were especially successful (Figure 3). Ground and local participant controllers issued approximately 72% of verbal instructions before or at the same time as ETCs,  $t(19) = -6.74, p < .01$ . However, local controllers issued many more instructions before ETCs than did ground,  $t(9) = 3.30, p < .01$ , while ground controllers issued more instructions at the same time,  $t(9) = 5.83, p < .001$ . Despite issuing instructions earlier, video data also revealed that the instructions issued by participants and ETCs were almost always the same, thus indicating that participants were still cognitively performing similar flight management processes as the control group.

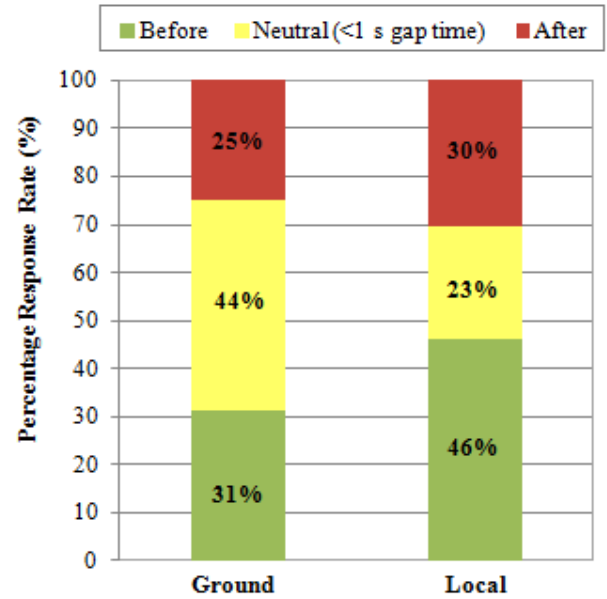


Figure 3: Percent response rate of verbal instructions. Chart legend presents gap time for each response

To analyze individual participant experiences and gather TFDM design suggestions, response rate and gap time were plotted together and video playback was used to revisit the circumstantial causes of when participants issued commands after ETCs. Only gap times over 2 s were further analyzed to eliminate noise due to timing precision. For example, in Figure 4 ground controller #6 clearly exhibited signs of high workload starting around 15:25 to which he began a slight acclimation toward the end of the run. Using video playback, it was discovered that he issued three instructions after the ETC because he was busy highlighting and editing a flight strip field on the FDM. This area is marked by three asterisks in Figure 4. Instances such as these were prevalent and examining them via video provided indications of potential design flaws to target for future system improvement.

Video data revealed that, for ground participant controllers, there was a clear difficulty in issuing timely instructions when searching for, moving, and editing electronic flight

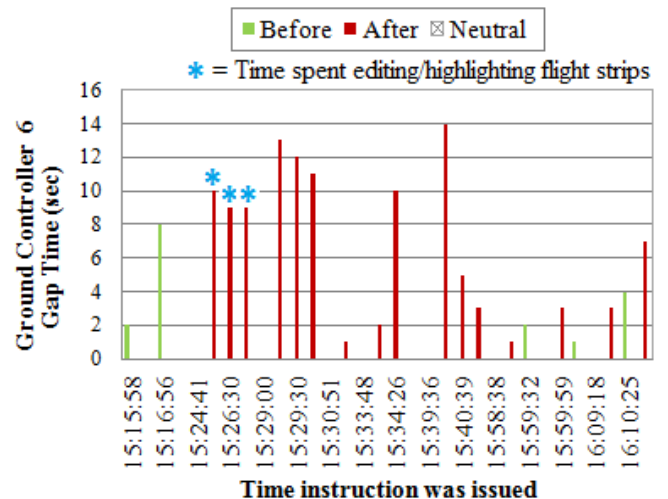


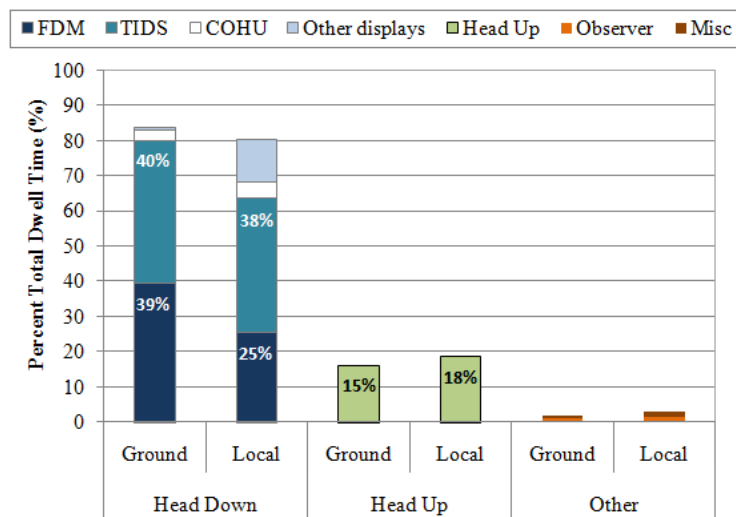
Figure 4: Sample response rate vs. gap time plots

strips on the FDM (23/54 instances). Local participant controllers also experienced this issue, though to a smaller degree (8/37 instances). Participants also had difficulty finding flight strips on the FDM due to an unreliable feature that automatically moved flight strips to a “Ready to Taxi” bin from their primary “Holding” bin. Due to the prototype nature of the system, this feature was only intermittently reliable and thus caused confusion about the location of new flight strips. Such findings indicate the need to revise the design of Edit and Search capabilities for flight strips on the FDM.

Video analyses also revealed that participants were sometimes prompted to issue instructions by hearing their ETC counterparts (ground: 7/54, local: 12/37 instances). This finding indicates that participants were clearly influenced by their ability to hear the ETC control group. While this is an innate limitation of the method, it does not appear to skew our results significantly: the less than 1 s gap in the majority of these cases provides evidence that participants were not simply repeating ETC clearances, but were prompted to execute their own formulated clearances upon hearing ETCs.

### Visual attention results

Visual gaze data were analyzed with paired t-tests to compare ground and local participant view rates of gaze areas. Gaze data were not collected for ETCs due to restrictions on data gathering from active controllers. Figure 5 provides a breakdown of percent total dwell time, with gaze areas listed in the key. The “Misc” and “Observer” categories are combined on the y-axis to create an “Other” category, and were created to separate noise from actual data. Respectively, they refer to instances when participants viewed non-informational areas and when they were looking at the observer.



**Figure 5: Percent total dwell time by controller type and dwell area.**

On average, participant controllers spent significantly more time in a head-down position ( $M = 81.9$ ,  $SD = 12.8$ ) than head-up ( $M = 17.2$ ,  $SD = 12.4$ ),  $t(43) = 17.12$ ,  $p < .001$ . They also spent more time looking head-up out the window over “Other” dwell areas ( $M = 2.3$ ,  $SD = 1.9$ ),  $t(43) = -7.87$ ,  $p <$

$.001$ . A significant amount of attention was clearly directed toward displays and information within the tower as opposed to outside of it. The top three areas particularly focused on by both ground and local controllers were the TIDS, FDM, and head-up areas. Total dwell times revealed a statistically significant difference between ground and local participants’ viewing of the FDM,  $t(20) = -5.36$ ,  $p < .001$ . This is an expected result because ground controllers spend more time editing and sequencing electronic flight strips (functions on the FDM).

Video playback was also referenced to find causes of the longest individual dwells. Categorizing the dwell area and perceived causes of all dwells over 10 s revealed that dwells between 10 and 15 s showed no new information added. Thus, only dwells over 15 s were further analyzed via video playback. Notably, the rate of dwells over 10 s was twice as high for ground (200 instances) than local (107), controllers. Since a large majority of these dwells focused on the TIDS and FDM, this suggests that ground controllers had more difficulty discerning information from the two displays. Upon investigating the causes of dwells over 15 s, ground controllers were found frequently editing electronic flight strips on the FDM (13/53), viewing/monitoring the TIDS (11/53), and using the COHU™ camera view on the TIDS (11/53). There were also multiple instances in which ground controllers had difficulty finding flight strips on the FDM and utilized the “Search” function. Only 8 out of 53 dwells occurred due to experimental noise, such as controller/observer interaction. For local controllers, a majority of long dwells occurred when monitoring or editing flight strips on the FDM (5/22), viewing/monitoring the TIDS (8/22), and using a COHU™ view within the TIDS (7/22). Only 2/22 dwells were due to noise-related causes.

In summary, visual attention analyses revealed that ground controllers heavily used flight strip features on the FDM, and both ground and local participants were found to be viewing/monitoring elements on the TIDS or monitoring a plane through the COHU™ camera view (in some cases this was requested by test personnel). While it is clear that controller interaction with flight strips on the FDM needs to be revised, it is questionable whether attention to the TIDS and COHU™ view in TIDS was a negative occurrence. While time spent head-down took participant attention away from live events outside the tower, both of these windows enabled continuous monitoring of flights on the runway. In addition, post-demonstration survey results indicated that participants felt the TIDS display highly advantageous to their work activities.

## DISCUSSION

In 1992, Bentley et al. used sociological observations of controller behavior to aid in the design of a flight management system. However, they note there was no clear mapping from their observations to specific design requirements or implementation. Comparatively, the non-intrusive measures of TFDM video and audio analyses aided in both quantifying data and providing a clear mapping of results to design changes. During field demonstrations, participants provided over 200 individual suggestions via surveys to communicate their

opinions and give design suggestions. Without performance assessment measures and video playback analyses to prioritize these results, it would have been difficult to understand the benefits of a change and thus its usefulness in being implemented (e.g., Shanks, Rowland, & Ranger, 2005; Frensch & Miner, 1994).TFDM video/audio analyses revealed areas of system improvement and also highlighted its successes. While changes to FDM search capability and electronic flight strip functions were justified through high dwell rates during these actions, a high visual attention towards TIDS coupled with survey results demonstrated its potential usability. Table 1 highlights actions causing high workload and high visual attention. These results pinpointed design and workload issues and provided considerations that can be used to inform the future design of TFDM and other ATC systems.

**Table 1: Circumstances/actions observed during high verbal communication gap times and visual dwells over 15 s.**

Issues Discovered Through:	Verbal Instructions	Visual Attention
Manually searching for flight strips on FDM	X	X
Editing & updating flight strips	X	X
Moving flight strips	X	
Difficulty using "Search" function	X	
Tracking flight on TIDS		X
Forgetting to update a flight strip		X
Slow when using FDM keyboard		X
Adjusting TIDS COHU™ view		X
Inconclusive cause	X	X

While participant controllers spent a significant amount of time head-down, results are comparable to previous studies on controller visual attention (e.g., Pinska, 2007; SMI, 2011). The relatively higher head-down rate seen in TFDM may not necessarily be caused by the system itself and instead may reflect participant adjustment to the new system and to researchers instructing participants to concentrate on system functionality during the demonstration.

In summary, concrete quantifications of visual attention and cognitive workload enabled researchers to find the most critical TFDM prototype flaws to improve upon in future system design iterations. The methods described in this paper also contribute to a more general battery of field-accepted methods of workload and attention assessment. Future replications may result in even stronger trends if sources of noise, such as controller/observer interaction and ETC influence upon participant commands, are addressed.

## ACKNOWLEDGEMENTS

The authors wish to thank the TFDM researchers and developers, especially Lianna Hall and Dan Tennant, who were responsible for video and audio data collection and compilation. This work was sponsored by the Federal Aviation Administration under Air Force Contract No. FA8721-05-C-0002. Opinions, interpretations, conclusions, and recommendations are those of the authors and are not necessarily endorsed by the United States Government.

## REFERENCES

- Averty, P., Collet, C., Dittmar, A., Athenes, S., & Vernet-Maury, E. (2004). Mental workload in air traffic control: An index constructed from field tests. *Aviation, Space, and Environmental Medicine*, 75(4), 333-341.
- Bentley, R., Hughes, J.A., Randall, D., Rodden, T., Sawyer, P., Shapiro, D., & Sommerville, I. (1992). Ethnographically-informed system design for air traffic control. *Proceedings of the Conference on Computer Supported Cooperative Work*. (pp.123-129). Toronto, Canada: Association for Computing Machinery.
- Berka, C., Levendowski, D.J., Lumicao, M.N., Yau, A., Davis, G., Zivkovic, V.T. ... Craven, P.L. (2007). EEG correlates of task engagement and mental workload in vigilance, learning, and memory tasks. *Aviation, Space, and Environmental Medicine*, 75(5), B231-44.
- Cardosi, K. (1993). *An analysis of en route controller-pilot voice communications*. (DOT Report DOT/FAA/RD-93/11). Washington, DC: FAA.
- Cardosi, K., Yost, A. (2001). Controller and pilot error in airport operations: A review of previous research and analysis of safety data. (DOT Report No. DOT/FAA/AR-00/51). Washington, DC: Federal Aviation Administration, Office of Aviation Research. (NTIS No. PB2001-104596).
- Davison Reynolds, H.J., Kuffner, M.P., & Yenson, S.K. (2011) A field demonstration of the air traffic control Tower Flight Data Manager prototype. *Proceedings of the Human Factors and Ergonomics Society 55<sup>th</sup> Annual Meeting*: 55(61), 61-65. doi: 10.1177/1071181311551013
- Embrey, D., Blackett, C., Marsden, P., Peachey, J. (2006, July). Development of a human cognitive workload assessment tool. *Human Reliability Associates*. Retrieved August 2011, from United Kingdom Department of Transport, The Maritime and Coastguard Agency website: [http://www.dft.gov.uk/mca/research\\_report\\_546.pdf](http://www.dft.gov.uk/mca/research_report_546.pdf)
- Endsley, M.R., & Rodgers, M.D. (1997). Distribution of attention, situation awareness, and workload in a passive air traffic control task: Implications for operational errors and automation. (DOT/FAA/AM-97/13). Washington, DC: FAA, Office of Aviation Medicine.
- Federal Aviation Administration. (2012). *NextGen implementation plan*. Retrieved from [http://www.faa.gov/nextgen/implementation/media/NextGen\\_Implementation\\_Plan\\_2012.pdf](http://www.faa.gov/nextgen/implementation/media/NextGen_Implementation_Plan_2012.pdf)
- Frensch, P.A., & Miner, C.S. (1994). Effects of presentation rate and individual differences in short-term memory capacity on an indirect measure of serial learning. *Memory & Cognition*, 22(1), 95-110.
- Funke, G.J., Knott, B.A., Salas, E., Pavlas, D., & Strang, A.J. (2012). Conceptualization and Measurement of Team Workload: A Critical Need. *Human Factors*, 54(1), 36-51.
- Howard, G.S., & Dailey, P.R. (1979). Response-shift bias: A source of contamination of self-report measures. *Journal of Applied Psychology*, 64(2), 144-150.
- Metzger, U. & Parasuraman, R. (2001). The role of the air traffic controller in future air traffic management: An empirical study of active control versus passive monitoring. *Human Factors*, 43(4), 519-528.
- Neale, J. M., & Leibert, R. M. (1973). *Science and behavior: An introduction to methods of research*. Englewood Cliffs, N.J.: Prentice-Hall.
- Parasuraman, R., & Riley, V. (1997). Humans and automation: Use, misuse, disuse, abuse. *Human Factors*, 39(2), 230-253.
- Pavet, D. Garron, & J. Coullon, I. (2001). Use of paper strips by Tower ATCO and promises offered by new design techniques on user interface. *4<sup>th</sup> USA/Europe ATM R&D Seminar*. Santa-Fe, NM: Federal Aviation Administration & EUROCONTROL. Retrieved from [www.atmseminar.org/seminarContent/seminar4/papers/p\\_087\\_HF.pdf](http://www.atmseminar.org/seminarContent/seminar4/papers/p_087_HF.pdf)
- Pinska, E. (2006). An Investigation of the Head-up Time at Tower and Ground Control Positions. *Proceedings of the 5<sup>th</sup> EUROCONTROL Innovative Workshop and Exhibition* (pp. 81-86). Bretigny sur Orge, France: EUROCONTROL.
- SensoMotoric Instruments (SMI). (2011). Case Study Eye Tracking: Remote Air Traffic Control. SensoMotoric Instruments GmbH. Retrieved from [http://www.smivision.com/fileadmin/user\\_upload/downloads/case\\_studies/smi\\_cs\\_atc.pdf](http://www.smivision.com/fileadmin/user_upload/downloads/case_studies/smi_cs_atc.pdf)
- Shanks, D.R., Rowland, L.A., Ranger, M.S. (2005). Attentional load and implicit sequence learning. *Psychological Research*, 69(5-6), 369-382.
- Wierwille, W.W., & Eggemeier, F.T. (1993). Recommendations for mental workload measurement in a test and evaluation environment. *Human Factors*, 35(2), 263-281.
- Wickens, C. D. (1992). *Engineering Psychology and Human Performance*, 2<sup>nd</sup> edition. HarperCollins Publishers, Inc.
- Wilson, G.F. & Russell, C.A. (2003). Operator functional state classification using multiple psychophysiological features in air traffic control task. *Human Factors*, 45(3), 381-89.