Multifactor interactions and the air traffic controller: The interaction of situation awareness and workload in association with automation

T. Edwards*, J. Homola, J. Mercer***, L. Claudatos******
* San Jose State University/NASA Ames Research Center, Moffett Field, CA 94035
USA (Tel: 650-604-2165; e-mail: tamsyn.e.edwards@nasa.gov)
** NASA Ames Research Center, Moffett Field, CA 94035
USA (e-mail: jeffrey.r.homola@nasa.gov)
*** NASA Ames Research Center, Moffett Field, CA 94035
USA (e-mail: joey.mercer@nasa.gov)
**** San Jose State University/NASA Ames Research Center, Moffett Field, CA 94035
USA (e-mail: lauren.e.claudatos@nasa.gov)

Abstract: Air traffic controllers (ATCOs) must maintain a consistently high level of human performance in order to maintain flight safety and efficiency. In current control environments, performance-influencing factors such as workload, fatigue and situation awareness can co-occur and interact to affect performance. However, multifactor influences and the association with performance are under-researched. This study utilized a high fidelity human in the loop en-route air traffic control simulation to investigate the relationship between workload, situation awareness and ATCO performance. The current study aimed to replicate Edwards, Sharples, Wilson and Kirwan's (2012) previous research, and extend the study by using a ex-controllers as participants, and comparing multifactor relationships across four levels of automation. Results suggest that workload and situation awareness may interact to produce a compound impact on controller performance. In addition, the effect of the interaction on performance may be dependent on the context and level of automation. Findings have implications for human-automation teaming in air traffic control, and the potential prediction, and therefore support, of ATCO performance.

Keywords: Air traffic control, laboratory simulation, multifactor interactions, automation, human performance, workload, situation awareness

1. INTRODUCTION

Air traffic controllers are responsible for the safety and efficiency of all air traffic. It is essential that controllers maintain a consistently high standard of human performance in order to maintain flight safety. Knowledge of the impact of human factors on controller performance is critical to understand and mitigate threats to performance.

Previous research has largely focused on the association between single factors (e.g., workload, fatigue) and performance, which has resulted in a comprehensive understanding of single factor influences (e.g., Reason, 1990). In current control environments however, the residual threats for incidents often result from the interaction of multiple human factors and the associated cumulative impact on performance. There is therefore a gap in understanding of multifactor interactions and the association with human performance.

In an attempt to address this gap, Edwards, Sharples, Wilson and Kirwan (2012) investigated multiple factor interactions and associations with controller performance within the framework of a 'human performance envelope'; rather than focusing on one individual factor (e.g. fatigue), the envelope framework considers a range of factors and how they collectively influence performance. Using a simulated air traffic control task, covariate performance-influencing factors, including workload, SA, and fatigue were measured using self-report scales. Results indicated that several factors known to affect controller performance do co-vary, and that factors appeared to interact to produce a compound impact on both safety and efficiency performance measures.

However, there were several limitations to Edwards et al.'s (2012) study that potentially limited the valid generalisation of results to an operational environment. First, participants were college students rather than air traffic controllers. Although participants received basic ATC training, trained controllers may perform very differently, potentially affecting the identified interaction relationships. In addition, simulation fidelity was low, potentially reducing valid generalisation of results. Therefore, "future research should replicate these results using a full-scale simulation with trained ATCOs as participants" (Edwards et al., 2012, p8). In addition, participants completed all air traffic tasks with minimal automation. With the increasing amount of automation in air traffic control, the identified interactions and associations with performance may not accurately represent factor associations and performance in the presence of increased automation.

The current study attempted to address these limitations as well as extend Edwards et al.'s (2012) previous work, by examining factor interactions across varying automated functions, using retired ATCOs as participants in a high fidelity simulation. The aims of this study were therefore to address the limitations of previous research, and support future research, by investigating multifactor interactions and the association with controller performance, and to extend Edwards et al.'s (2012) research by investigating these effects in the context of different levels of available automation.

2. METHOD

The study reported in this paper is part of a larger study. Only the measures that are relevant to this paper are presented. An en-route air traffic control (ATC) human in the loop (HITL) simulation was utilised to investigate the relationship between workload and situation awareness (SA), and the subsequent association with performance. Workload and SA were selected for investigation as these factors have previously been identified as critical factors that frequently negatively influence controller performance (Edwards et al., 2012). Performance was inferred from the time taken to detect pre-programmed conflicts. Participants were eight ex-ATCOs who had worked in enroute airspace in Oakland Air Route Traffic Control Centre (ARTCC). Pseudo pilots with an average of over 500 logged in hours on the MACS pilot platform were paired with the controllers.

The study used a within measures design. The simulation was centred on high altitude, en-route airspace from the Cleveland ARTCC. Each participant was assigned to work sector 79, which was observed from previous HITL simulations to be a particularly complex sector given the mix of traffic transiting its airspace. Winds for the area were included, which were constant at altitude with forecast error.

Each simulation session lasted for 60 minutes. Traffic was approximately 20 percent more than the current day maximum traffic of 18 aircraft in the sector. The taskload level was created by changing the number of aircraft in the controlled sector and the complexity of the task by the number of aircraft requiring vertical movements and the number of aircraft pairs set on a conflicting flight path (Brookings, Wilson, & Swain 1996). The scenarios were designed to have ramp-up, sustained, and ramp-down phases, with each phase lasting approximately 20 minutes. The traffic was a mixture of overflights at level altitude, area arrivals with a top of descent in or near the simulation sector, and area departure aircraft that resulted in aircraft climbing from their origin airport into sector 79.

To investigate factor interactions and associated performance in accordance with different automation capabilities, controllers participated in four conditions: Condition 1- Conflict Detection (CD) where the participant was only responsible for detecting conflicts while all other tasks were automated, Condition 2 - Conflict Detection + Routine Tasks (RT) where the participant was responsible for detecting conflicts, accepting aircraft hand-offs from adjacent sectors with pilot check-ins, and initiating hand-offs as aircraft exit the test sector and issuing a frequency change, Condition 3 - Conflict Detection + Decision Making (DM) where the participant was responsible for conflict detection and making decisions based on fielded requests from flight crews and coordination with adjacent sectors, and Condition 4 - Conflict Detection + Routine Tasks + Decision Making where the participant was responsible for all tasks covered in the previous three tasks. Conditions were counterbalanced. To measure conflict detection performance, controllers were asked to enter a keyboard command to identify when they perceived an aircraft pair was in conflict.

2.1 Participants

A total of 8 participants (1 female, 7 male) took part in the simulation. Age ranged from 50 years – 69 years. Participants responded to grouped age ranges and so an average age could not be calculated. All participants were ex-controllers. Participants had worked as en-route controllers in the Oakland ARTCC. Years of experience as an ATCO (excluding training) ranged from 23 – 29.5 ($M=24.94$, $SD=2.54$).

2.2 Measures

In line with Edwards' (2012) previous study, covariate factors were measured using subjective, self-report scales. Mental workload was measured using the uni-dimensional Instantaneous Self-Assessment scale (ISA) (Tattersall & Foord, 1996). SA was measured using the Situation Present Assessment Method (SPAM) (Durso et al., 1995). Every 3 minutes, participants were presented with the ISA rating scale at the top of the radar scope and asked to click on the workload rating. After responding to the workload scale, a SA question (e.g., 'Will more than 3 aircraft leave your sector in the next minute?') was presented at the top of the simulation screen. Questions used a binary response format. Questions were developed in collaboration with 3 ex-ATCOs.

Several performance measures were collected during the simulation. Due to space constraints, only one of these performance variables will be examined in this paper: time to accurately detect conflicts. This variable was selected due to the important safety implications of this performance measure. In addition, in contrast to measures such as number of conflicts accurately detected, this measure allows for greater granularity in performance measurement and is not affected by potential ceiling effects. A conflict was defined as aircraft that would breach the separation minima without intervention. An internal conflict probe was used to identify conflicts and assess participants' conflict detection performance. Time to detect conflicts was determined from the time at which the data tag of both aircraft in conflict were first visible to the controller, subtracted from the time the controller made the identification keyboard entry. Measures were recorded continuously in the simulation software.

2.3 Apparatus

The software used was the Multi-Aircraft Control System (MACS) (Prevot et al., 2010). Participant workstations were configured with a BARCO large-format display and specialized keyboard/trackball combination that is representative of what is currently used in air traffic control facilities. Voice communications were enabled via a custom, stand-alone system that is also representative of what is used in operations. Data were collected continuously through MACS's data collection processes.

3. RESULTS

Due to the quantity of analyses and results, only results for the strongest and most relevant data trends will be presented in this article.

3.1 Workload and automated conditions

Workload was reported to be higher, on average, in the least automated conditions compared to the most automated conditions (Table 1), indicating variation in reported workload in association with number of automated functions.

Table 1. Average workload ratings by condition

Condition	Mean	SD
(1) Conflict Detection (CD)	3.09	0.54
(2) CD + Routine Tasks (RT)	3.46	0.74
(3) CD + Decision Making (DM)	3.39	0.46
(4) CD + RT + DM	3.71	0.76

A factorial repeated measures analysis of variance (ANOVA) confirmed a significant main effect of condition on workload ratings, $F(3, 21) = 8.74$, $p<0.005$. Pairwise comparisons revealed that workload ratings were significantly lower in the most automated condition (CD only) than CD + Routine Tasks ($p<0.05$), CD + Decision Making ($p<0.01$) and CD + RT + DM ($p<0.005$). Workload ratings in the CD + Routine Tasks condition were significantly lower than the least automated condition, CD + RT + DM ($p<0.05$). Finally, workload ratings in the CD + Decision Making condition were significantly lower than the least automated condition, CD + RT + DM ($p<0.05$). No other significant differences were found between conditions.

3.2 Situation awareness and automated conditions

On average, participants responded to SA questions more slowly in the most automated condition (condition 1, CD only) and condition 3, CD + DM (Figure 1), suggesting reduced SA. Participants appeared to respond faster in condition 2 (CD + RT) and the least automated condition (condition 4, CD + RT + DM), suggesting good SA. A factorial repeated measures analysis of variance (ANOVA) confirmed a significant main effect of condition on average response times to SA questions, $F(3,21) = 9.37$, $p<0.001$. Pairwise comparisons revealed that responses to SA questions were significantly slower in the most automated condition compared to the CD + RT ($p<0.01$), and the least automated condition ($p<0.05$). On average, responses to SA questions in the CD + RT were significantly faster than CD + DM ($p<0.005$). Finally, on average, responses to SA questions in the CD + DM condition were significantly slower than the least automated condition (CD + RT + DM) ($p<0.005$).

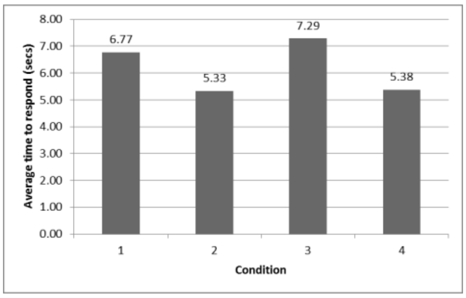

Fig. 1. Average time in seconds to respond accurately to SA questions

3.3 Performance across conditions

Time to detect conflicts appears to be slowest in the most automated condition (M=340.39, SD=39.30). Time to detect conflicts was fastest in the second condition, CD + routine tasks (M=282.06, SD=64.32) (Figure 2).

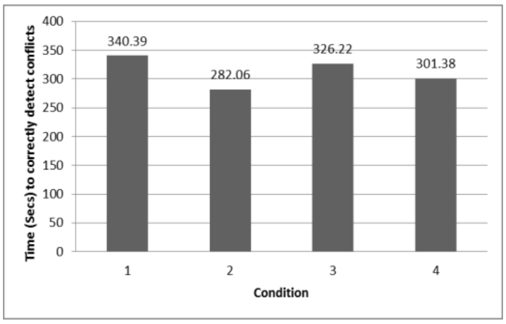

Fig. 2. Time in seconds to correctly detect conflicts per automation condition

A factorial repeated measures ANOVA was conducted to further examine the effect of condition on time to detect conflicts. There was a significant main effect of condition on time to detect conflicts, $F(3,21)=4.62$, $p<0.05$. Pairwise comparisons confirmed that participants detected conflicts significantly faster in condition 2, CD + RT, compared to condition1 - CD only ($p<0.005$), condition 3 - CD + DM ($p<0.05$) and condition 4 - CD + RT + DM ($p<0.05$). No other significant differences were found between conditions.

3.4 Does covariance between workload and SA exist?

Workload was expected to significantly negatively correlate with SA. Figure 3 presents a scatterplot of the association between workload and SA for condition 1 (other scatterplots are not included for brevity). A possible curvilinear relationship can be observed, a trend that was observed in each condition. There appears to be a relationship between increasing workload and increasing time to respond to SA questions until workload rating 4. At workload rating 5, time to respond decreases. No significant correlation was found between workload and SA for any condition. The relationship in condition 4, the least automated condition, approached significance (*rs*=0.15, $p<0.1$).

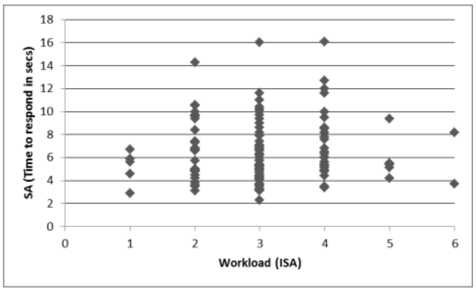

Fig. 3. Scatterplot of SA against workload (ISA scale is five-point, i.e. 1-5, 5 being very high) in condition 1, CD only

3.5 The interaction between workload and SA – is there a compound effect on performance?

The analysis was extended to investigate interactions between workload and SA, and the association with time to detect conflicts, across automation conditions. A median split approach was utilized (Denollet et al., 1996) to transform the continuous data into discrete factor groups. The following section presents findings from this analysis. No significant differences between groups were found. However, recurrent trends were identified that are worthy of further consideration.

Time to correctly detect conflicts was slowest in condition 1 (CD only) when low workload and poor SA (inferred from slower response times to SA questions) were combined (M=323.97, SD=53.29) (Figure 4). A trend of note is that both slowest times to detect conflicts occurred under low workload. Times to detect conflicts under low workload were faster when SA was good (inferred from faster accurate responses to SA questions), suggesting that good SA may support performance under a low workload, compared to poorer SA. This may be a compound effect – the time to detect conflicts increased disproportionately when low workload and poor SA were combined.

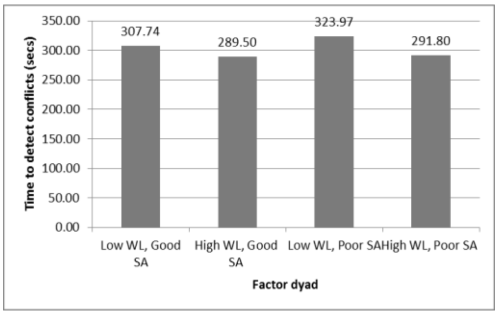

Fig. 4. Time in seconds to correctly detect conflicts in condition 1 – Conflict detection only

The trend seen in condition 2 is different to condition 1 (CD+ RT) (Figure 5). In condition 2, high workload, when combined with poor SA, appears be associated with slower times to detect conflicts compared to low workload. When high workload and poor SA co-occurred, time to detect conflicts was slower than in any other condition. The combination of high workload and poor SA may have interacted to be associated with a compound influence on performance.

A trend of note is that performance appears to remain stable in association with different multifactor combinations. The co-occurrence of one factor from the group associated with fastest time to detect conflicts (i.e. low workload or good SA) and one factor from the group associated with slowest time to detect conflicts (i.e. high workload or poor SA) were associated with very similar times to detect conflict.

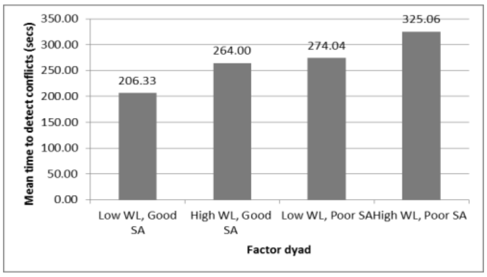

Fig. 5. Time in seconds to correctly detect conflicts in condition 2 – Conflict detection and routine tasks

The data shown in Figure 6 surprisingly suggests that in condition 3 (CD + DM) time to detect conflicts was fastest under a low workload condition, but combined with slower responses to SA questions. The slowest time to detect conflicts on average appears to be a combination of a low workload and faster responses to SA.

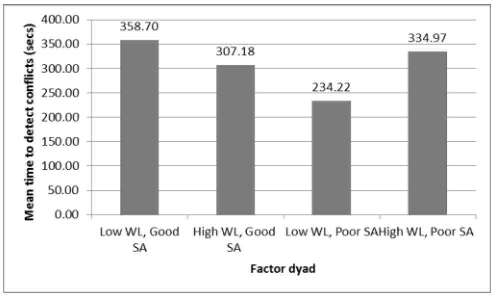

Fig. 6. Time in seconds to correctly detect conflicts in condition 3 – Conflict detection and decision making

Finally, in the least automated condition (condition 4, CD + RT + DM), participants only fell into one of two groups: low workload and good SA or high workload and poorer SA (Figure 7). This suggests that the demand may have influenced ATCO's ability to maintain a complete picture. Similar to condition 2 (CD + RT), low workload and fast SA response times appear to have been associated with faster detection of conflicts.

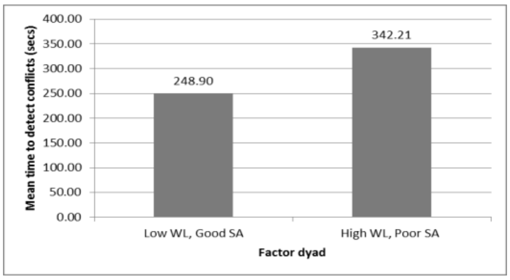

Fig. 7. Time in seconds to correctly detect conflicts in condition 4 – Conflict detection, routine tasks and decision making

4. DISCUSSION AND CONCLUSIONS

4.1 Workload, situation awareness and performance across automation conditions

Workload was found to be rated significantly differently between automation conditions as expected, with workload in condition 1 (most automated condition) rated significantly lower than condition 4 (least automated condition). The same trend was not seen with SA. As per the SPAM method (Durso et al., 1995), SA was inferred from the time taken to answer SA related questions correctly. Using this measure, faster response times (indicating better SA) were identified in conditions 2 (CD + RT) and the least automated condition 4 (CD + RT + DM). One explanation of these results is that the low workload in the most automated condition may have been experienced by participants as underload, and resulted in a reduction of engagement with the simulation, leading to reduced SA (e.g., Endsley 1999). The routine tasks of accepting and handing off aircraft (condition 2), and conducting tasks with minimal automation (condition 4) may have enabled the controllers to update their picture sufficiently to maintain a high level of SA. The safety related performance measure of time to correctly detect conflicts also varied significantly between conditions. Time to detect conflicts, on average, was slowest in the most automated condition (condition 1), fastest in condition 2 (CD+RT), and second fastest in the least automated condition (condition 4). It is important to note that this pattern of results is the same as the pattern identified in the SA measure, suggesting that, as would be expected, SA and time to detect conflicts are related. One potential explanation for the slower time to detect conflicts in the most automated condition is, as discussed previously, a low workload or underload may have negatively affected SA, ultimately influencing time to detect conflicts. If this is the case, this result has an obvious safety implication; automated systems should support the operator in maintaining situation awareness. An alternative explanation, however, may be that controllers purposefully left the situation to develop longer in the most automated condition before confirming that aircraft were in conflict. Controllers use a strategy of letting the situation develop longer when they have sufficient spare capacity to recover the situation if required (Edwards et al., *In Press*). This strategy ensures the most efficient performance; controllers can reduce the number of unnecessary changes to an aircraft by waiting to see if aircraft pairs definitely will be in conflict. However, under higher workloads, controllers issue instructions sooner in order to keep ahead of the traffic (Edwards et al., *In Press*) which may be reflected in the time to detect conflicts.

4.2 Covariance between workload and SA

For all conditions, the relationship between workload and SA was unexpectedly weak, replicating the findings of Edwards et al.'s (2012) previous study. The weak relationship found in this study may be due to a fallacy of the linear correlation analysis that was applied. It appears that a curvilinear relationship may exist between the two measures (Figure 3) in which time to respond to SA questions increased (indicating poorer SA) with workload until point '4' on the ISA scale. However, as workload continues to increase, time to respond becomes faster. This may be explained that when controllers experience these high levels of workload, they respond to the SA question as quickly as possible to remove it from the scope without necessarily paying full attention to the accuracy of the answer.

4.3 The interaction between workload and SA and the association with performance

Although significant differences between factor dyad groupings were not found, the recurrent observed data pattern is important to consider. There may be several reasons for a lack of statistical significance. Specifically, the method of median splits has been criticised for creating a loss of variance in the data, therefore reducing the power of applied statistical tests (Maccallum, Zhang, Preacher & Rucker, 2002) and increasing the risk of type II error. In addition, the relatively small sample of eight participants may have affected statistical power. However, recurrent data trends suggest robustness even without statistical significance, and are therefore considered to be important for discussion.

A robust data trend was identified where specific factor combinations were associated with greater performance changes than when the factors occurred independently in other factor groups. This result was reported across conditions. An interpretation of this finding is that co-occurring factors may interact and are associated with a compound influence on performance. Performance changes in association with factor groups may be positive or negative; combinations of factors are associated with increased performance, greater than performance achieved in association with the factors occurring independently, as well as performance declines.

An additional data trend of note is that average performance measures can remain stable when associated with different co-occurring factors. An example of this data trend was identified in condition 2. Low workload and good SA were associated with the fastest time to detect conflicts out of the four factors groupings. This may be expected, as a lower workload (note: not underload) may support development and maintenance of SA. When these factors independently occurred with other factors (e.g. low workload and poor SA; high workload and good SA) performance was similar. One interpretation of this finding is that when factors are combined, a factor associated with increased performance may moderate a factor that is associated with performance declines. Both of these data trends were also identified in Edwards et al.'s (2012) previous study, confirming that the previously identified data trends appear to be robust across participant groups and differing levels of automation availability.

The findings of the present study extend Edwards et al.'s (2012) previous work by considering the interaction relationship across differing levels of automated functions. The association of the factor groups and performance appeared to differ depending on condition. For example, in condition 1 (most automated condition), high workload and good SA resulted in the fastest average conflict detection time. In contrast, in condition 2 (CD +RT), low workload and good SA resulted in the fastest average conflict detection time. It is therefore important to interpret results in context. As may be expected, a high workload may elicit a higher level of alertness in participants (Repetti, 1993). When controllers are monitoring only, a higher workload may support alertness and engagement, resulting in a faster time to detect conflicts. However, in conditions 2 and 4, the same level of taskload may now be experienced to be a higher workload than experienced in condition 1, negatively influencing performance. It is important to note that the same co-occurring factors may have different associations with performance depending on the control task. The data trends observed in condition 3 deviate from the data trends observed in the other conditions. One explanation may be that controllers were distracted by decision making tasks, prioritising decision making over answering the SA questions. If this is the case, it may reflect why controllers who rated experiencing a low workload and slower responses to SA questions appear to detect conflicts faster than other groups.

These findings have important implications for both practical applications and future research. If recurrent trends can be identified, ops-room supervisors may be able to implement strategies to support controller performance prior to a potential performance decline, preventing performance-related incidents. In addition, by confirming and extending Edwards et al.'s (2012) previous research, there appears to be growing support for the concept of co-occurring, multifactor interactions and the associations with performance. Further research should investigate additional multifactor relationships, and the association with various safety and efficiency measures of performance, towards the potential development of a model of factor co-occurrences and predicted associations with performance. Further research may also investigate if findings can be applied to adaptive automation research to provide predictive guidance regarding when automation may be used to support the controller most effectively.

REFERENCES

Brookings, J. B., Wilson, G. F., & Swain, C. R. (1996). Psychophysiological responses to changes in workload during simulated air traffic control. *Biological Psychology*, 42, 361-377.

Denollet, J., Rombouts, H., Gillebert, T. C., Brutsaert, D. L., Sys, S. U., and Stroobant, N. (1996). Personality as independent predictor of long- term mortality in patients with coronary heart disease. *The Lancet*, 347(8999), 417-421.

Durso. F. T. Truitt, T. R., Hackworth. C, Crutchfield. J., Nikolic. D., Moertl. P., et al. (1995). Expertise and chess: A pilot study comparing situation awareness methodologies. In D. J. Garland & M. R. Endsley (Eds.), *Experimental analysis and measurement of situation awareness* (pp. 295-304). Daytona Beach. FL: Embry-Riddle Aeronautical University Press,

Edwards, T., Sharples, S., Wilson, J. R., and Kirwan, B. (2012). The need for a multi-factorial approach to safe human performance in air traffic control. *Proceedings of the 4th AHFE International Conference,* 21-25th July, San Francisco: USA.

Edwards, T., Sharples, S., Kirwan, B., Wilson, J. R. (*In Press*) Identifying Markers of Performance Decline in Air Traffic Controllers. In Vallicelli, A., Stanton, N.A., Landry, S. J. (Eds.) *Human Factors in Transportation: Social and Technological Evolution Across Maritime, Road, Rail, and Aviation Domains*.

Endsley, M. R. (1999). Situation awareness in aviation systems. *Handbook of aviation human factors*, 257-276. CRC Press, FL, USA

MacCallum, R. C., Zhang, S., Preacher, K. J., & Rucker, D. D. (2002). On the practice of dichotomization of quantitative variables. *Psychological methods*, 7(1), 19.

Prevot, T., Lee, P., Callantine, T., Mercer, J., Homola, J., Smith, N. and Palmer, E., (2010). Human-in-the-Loop Evaluation of NextGen Concepts in the Airspace Operations Laboratory. In: *Proceedings of the AIAA Modeling and Simulation Technologies Conference*, AIAA, Toronto, Canada

Reason J. (1990). *Human Error*. New York: Cambridge University Press

Repetti, R. L. (1993). Short-term effects of occupational stressor on daily mood and health complaints. *Health and Psychology*, 12(2), 125- 131.

Tattersall, A.J., and Foord, P.S. (1996). An experimental evaluation of instantaneous self-assessment as a measure of workload. *Ergonomics*, 39(5), 740–748.

Wickens, C. D. (2002). Situation awareness and workload in aviation. *Current Directions in Psychological Science*, 11(4), 128-133.