TRACKING WORKLOAD AND ENGAGEMENT IN AIR TRAFFIC CONTROL STUDENTS USING ELECTROENCEPHALOGRAPHY COGNITIVE STATE METRICS

Kyle A. Bernhardt
University of North Dakota Department of Psychology
Grand Forks, North Dakota
Terra Jorgenson, Craig Carlson, Paul Drechsel, Colt Iseminger
University of North Dakota Department of Aviation
Grand Forks, North Dakota
Dmitri Poltavski, F. Richard Ferraro, Thomas Petros
University of North Dakota Department of Psychology
Grand Forks, North Dakota

The current study evaluated the utility of electroencephalography (EEG) cognitive state to track workload and engagement changes in air traffic control students of differing experience during a Terminal Radar Approach Control (TRACON) scenario. EEG recordings were collected from *47* air traffic control students (27 with high and 20 with low experience) during a five phase TRACON scenario. The scenario fluctuated in the number of aircraft released per phase and the presence or absence of uncontrolled departures/arrivals. EEG workload probabilities were higher during the phase with uncontrolled departures/arrivals and maximum number of aircraft compared to phases with no uncontrolled arrivals/departures and fewer aircraft. Metrics of engagement did not vary throughout the scenario. Trends toward experience level differences in EEG metrics were observed, with less experienced students displaying slightly higher workload and engagement probabilities compared to their more experienced counterparts. Both experience groups made the most errors after the highest workload period.

The use of psychophysiological measurements to monitor operator cognitive states in the workplace is central to the field of neuroergonomics. Over the past two decades, a significant amount of research in the applied sector has focused on psychophysiological measures (e.g., electroencephalography) to monitor operators and evaluate their cognitive state (Parasuraman, 2015). In the profession of air traffic control (ATC), workload and engagement are two cognitive states critically tied to performance (Desmond & Hoyes, 1996; Signal, Gander, Anderson, & Brash, 2009). The complexity of ATC operations fosters an environment where overload situations can rapidly occur, potentially impairing controller decision making and placing aircraft in dangerous situations (Wickens, Mavor, & McGee, 1997). Moreover, low engagement, underload situations may lead to performance decrements as well (Desmond & Hoyes, 1996; Hancock & Warm, 1989; Saxby, Matthews, Warm, Hitchcock, & Neubauer, 2013). Psychophysiological measures of these constructs have been proposed as an alternative to performance-based metrics for evaluating an operator's skill level. Harrison et al. (2014) showed cerebral hemodynamic changes in ATCs as they learned to utilize Next-Generation technology. Moreover, Johnson et al. (2014) demonstrated electroencephalographic (EEG) differences in expert and novice shooters during a deadly force shooting simulator, with experts showing greater frontal theta activation and overall alpha suppression compared to novices. Experts also showed lower levels of EEG metrics of engagement compared to novices. Experts and novices did not differ with respect to EEG metrics of workload; however, the group size for each skill level was small (n = 6).

EEG provides several advantages for operator state assessment such as low cost, ease of application (Parasuraman, 2015), and the availability of turn-key systems that offer validated metrics of workload and engagement (e.g., Berka et al., 2007). Thus, EEG provides a suitable means for assessing ATC skill acquisition in ATC training institutions in addition to performance metrics, which only provide a unidimensional assessment of operator skill. That is, although two individuals may show the same task performance, one individual may experience significantly more workload and elevated task engagement than the other. The former individual would then be less adept at handling emergency situations than the latter due to reduced resource allocation capacity (Wickens & Tsang, 2015). More training may be necessary to ensure successful adaptation to increased workplace demands. The current study investigated the utility of using EEG derived cognitive state metrics of workload and engagement to evaluate ATC students of differing experience levels during a dynamic Terminal Radar Approach Control (TRACON) scenario. It was predicted that less experienced students would exhibit higher workload and engagement as well as worse performance than more experienced students.

Method

Participants

A total of 49 undergraduate air traffic control training students (two women and 47 men) from the University of North Dakota Department of Air Traffic Control participated in this study (M_{age} = 20.97, SD_{age} = 1.01). At the time of the study, all participants were enrolled in air traffic control training courses. From their prior coursework and advanced standing, all participants had adequate and on-going experience with the general procedures used with the TRACON simulator; however, they remained naïve with respect to the scenario to be performed. Two participants were excluded because of simulator technical difficulty, resulting in a final sample size of 47. Participants were split into low (n = 20) and high (n = 27) experience levels based on the highest ATC course completed that allowed them to perform all aspects of the scenario adequately.

Simulator and Scenario

This study utilized a high-fidelity TRACON simulator running ATCoach Global (V.4.32.5) Air Traffic Software with a Ubuntu Linux 11.04 (32-bit) operating system. The simulated RADAR scope was presented on a computer monitor measuring 20.1 in diagonal viewable area (1600 x 1200 resolution) with a 4:3 aspect ratio. Controllers utilized a keyboard and trackball mouse during the scenario to accept/initiate aircraft handoffs and manipulate aircraft data tags. During the scenario, controllers gave verbal instructions (e.g., altitude changes) communicated via a headset connected to a local radio connection to two research assistants (graduates of the air traffic control program) designated as “pseudo-pilots” located out of view from the controller’s station. The pseudo-pilots utilized stations similar to the controller’s station and inputted the commands for the aircraft in accordance with the controller’s instructions. Pseudo-pilots gave verbal readbacks of controller commands.

This study utilized Academy Airspace, a fictitious training airspace used by the Federal Aviation Administration (FAA) to train new controllers. The Academy airspace consists of centrally located Academy Airport (KAAC) and five surrounding airports (one towered military airport and four uncontrolled airports). All inbound and outbound aircraft entered/departed the airspace through established arrival/departure gates. The scenario consisted of three types of aircraft: jets (both light and heavy), turboprop, and reciprocating. Participants gained control of inbound aircraft 40 nautical miles out from KAAC and were responsible for controlling the aircraft to their designated airport and clearing the aircraft for published instrument approaches (ILS, NDB, or GPS approach). All airports were operating under single runway operations.

Participants performed a single 1-hr and 15-min continuous scenario composed of five, 15-min phases. The phases were developed during prior work to create a multidimensional, variable task demand TRACON scenario. All simulated aircraft were operating under instrument flight rules (IFR). Arrivals and departures occurred at both KAAC and the five satellite airports. Phases were distinguished via two workload-contributing factors: (1) number of aircraft arriving or departing the airspace within a given 15-min phase and (2) the presence or absence of arriving and departing uncontrolled aircraft. Phases were designated 1, 2, 3, 4, 5 and consisted of 8, 11, 16, 8, and 3 departing/arriving aircraft, respectively. Uncontrolled aircraft occurred during Phase 3. From these phase parameters, it was predicted that workload and engagement would follow a negative quadratic function shape, with peak values occurring at Phase 3. Manipulations of air traffic volume have shown to reliably elicit workload changes in ATCs (Vogt, Hagemann, & Kastner, 2006). Moreover, handling uncontrolled aircraft strains working memory capacity. Working memory strain has also been shown to be a reliable manipulation of operator workload (Wickens & Tsang, 2015). To add realism to the scenario, controllers were also required to complete flight strip marking procedures for each aircraft. Paper strips housed in plastic holders were located to the right of the controller in two bays for active and non-active aircraft.

Scenario scoring. Scenario runs were recorded by the TRACON simulator program and later scored by a team of University of North Dakota Department of Air Traffic Control faculty. Combining for over 40 years of experience, these faculty members have backgrounds in either controlling professionally or were graduates of the University of North Dakota Air Traffic Control program. Participant performance was evaluated by tracking the number of errors made in phraseology, violation of aircraft separation minimums, and airspace procedural violations for each 15-min phase. Errors were identified according to violations of standards set by FAA Order JO 7110.65W and letters of agreement (LOAs) established by Academy Airspace for training purposes.

EEG Recording

EEG was recorded using the Advanced Brain Monitoring (ABM) B-Alert X24 wireless Bluetooth system sampling at 256 Hz. The system incorporates 20 electrodes placed according to the international 10/20 system: Fp1, Fp2, F7, F3, Fz, F4, F8, T3, C3, Cz, C4, T4, T5, P3, Pz, P4, T6, O1, POz, and O2. Reference electrodes were placed at the left and right mastoids. Data were filtered with 50, 60, 100, and 120 Hz notch filters with Low Pass FIR filters online during data collection. Proprietary ABM artifact decontamination algorithms removed artifacts resulting from electromyography, eye blinks, excursions, saturations, and spikes. Power spectral density (PSD) was calculated on a second by second epoch frequency by applying a 50% overlapping Kaiser window for data smoothing to three data point windows consisting of 256 decontaminated data points each. These data were then subject to Fast Fourier Transformation resulting in four standard bandwidths (delta, alpha, theta, and beta). ABM proprietary algorithms (Berka et al., 2007) then computed cognitive state metric probabilities of high engagement and average workload ranging from 0.00 to 1.00 based on PSD values. Higher values indicate a higher probability of being in an engaged or overload state, respectively. EEG metrics were averaged within each 15-min phase using 5% trimmed means to eliminate extreme values.

Procedure

Participants first arrived at the TRACON simulation room on the University of North Dakota campus and provided written consent. Then, participants completed individual difference measures (results not reported here) and were fitted with the B-Alert EEG headset. To obtain ABM cognitive state metrics, participants performed three computerized cognitive benchmark tasks: 3-choice vigilance, visual psychomotor vigilance test, and an eyes-closed auditory psychomotor vigilance test. One of the trained pseudo-pilots then briefed the participant on the scenario and gave instructions regarding strip marking procedures. The simulation was then commenced and ran continuously for 1.25-hr.

Data Analytics

Series of 5 (scenario phase: 1, 2, 3, 4, 5) x 2 (experience: low, high) mixed model ANOVAs were conducted to evaluate scenario performance and EEG cognitive state metrics. Scenario phase served as the within-subjects factor and experience served as the between-subjects factor. Analyses were evaluated at an $\alpha = .05$. If the assumption of sphericity was violated, degrees of freedom corrections were employed. When ε values were less than .75, the Greenhouse-Geisser correction was used and when ε values were greater than .75, the Huynh-Feldt correction was used (Field, 2009). Sidak corrections were utilized for multiple comparisons if pairwise comparisons were warranted.

Results

Simulation Performance

Phraseology. Graphical representations of performance data are displayed in Figure 1. The results of the mixed model ANOVA for phraseology errors revealed a significant main effect for phase, $F(4, 180) = 19.72$, $p < .001$, $\eta_p^2 = .31$. Overall, the number of phraseology errors significantly increased from Phase 1 to Phase 2 and from Phase 3 to Phase 4. Phraseology errors then decreased during Phase 5, but not significantly from Phase 4 (see Figure 1). Neither the main effect for experience nor the experience by phase interaction were significant, $Fs < 1$, ps > .05.

Separation. Analysis of separation errors revealed a significant main effect for phase, $F(2.31, 103.75) = 43.70$, $p < .001$, $\eta_p^2 = .49$. Separation errors significantly increased after Phase 2 and reached a maximum at Phase 4, and then significantly decreased during Phase 5 (see Figure 1). The main effect for experience and the experience by phase interaction were not significant, $Fs < 1$, $ps > .05$.

Procedures. The results of a mixed ANOVA analyzing procedural errors revealed a significant main effect for phase, $F(2.71, 121.91) = 108.69$, $p < .001$, $\eta_p^2 = .71$. Similar to phraseology and separation errors, procedural errors significantly increased from Phase 1 onward until reaching a maximum at Phase 4. Procedural errors then decreased from Phase 4 to Phase 5 (see Figure 1). The main effect for experience, $F < 1$, $p > .05$, and the interaction between phase and experience, $F(2.71, 121.91) = 1.89$, $p = .14$, $\eta_p^2 = .04$, were not significant.

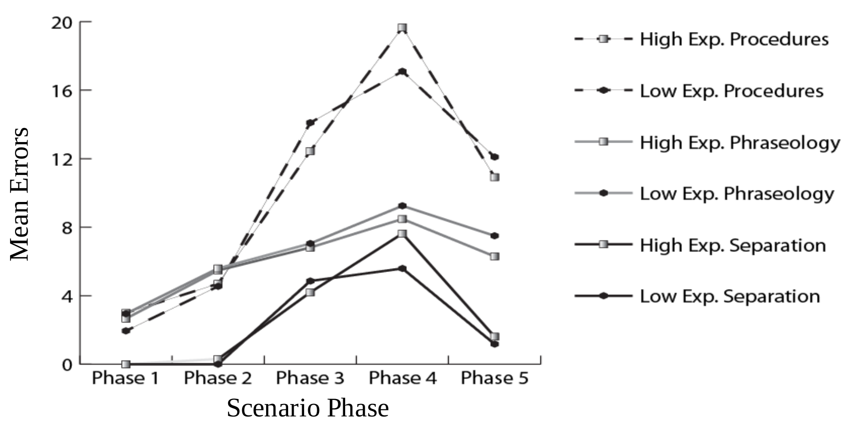

Figure 1. Scenario phase performance by controller experience (Exp.).

EEG Metrics

Workload. Results of EEG metrics are displayed in Figure 2. The results of a mixed ANOVA analyzing the B-Alert average workload metric revealed a main effect for phase, $F(2.29, 102.90) = 12.79$, $p < .001$, $\eta_p^2 = .22$. Examination of the group means demonstrated that workload increased across the phases, peaking at Phase 3 (see Figure 2 left panel). Phase 3 was significantly higher in workload than Phase 1, Phase 2, and Phase 5. Moreover, Phase 4 was significantly higher in workload than Phase 1 and Phase 5. The main effect for experience failed to reach significance, $F < 1$, $p > .05$. Additionally, the interaction between phase and experience was not significant, $F(2.29, 102.90) = 1.72$, $p = .18$, $\eta_p^2 = .04$.

Engagement. The analysis of the B-Alert high engagement metric (see Figure 2 right panel) revealed a nonsignificant main effect for phase, $F(2.62, 118.00) = 2.40$, $p = .80$, $\eta_p^2 = .05$, and a nonsignificant main effect for experience, $F(1, 45) = 2.59$, $p = .11$, $\eta_p^2 = .05$. The phase by experience interaction was also not significant, $F(2.26, 118.00) = 1.19$, $p = .31$, $\eta_p^2 = .03$.

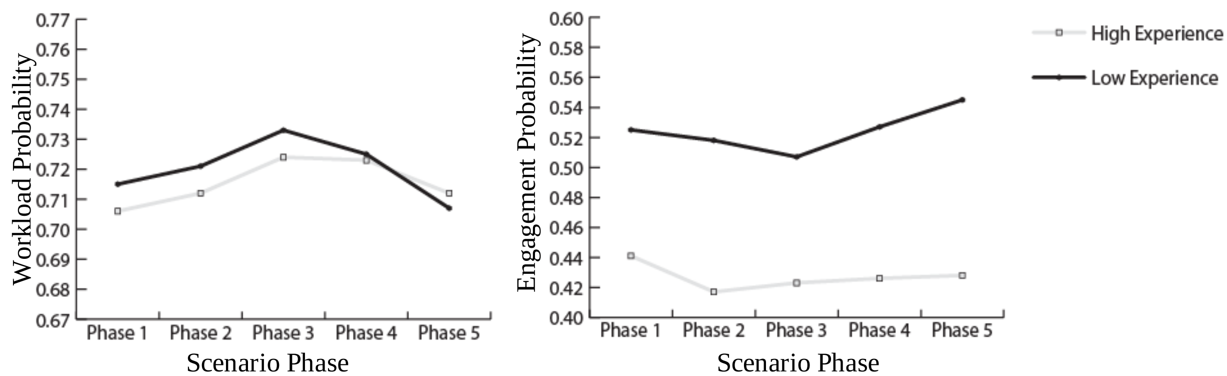

Figure 2. Scenario phase EEG workload (left panel) and engagement (right panel) probabilities by experience.

Discussion

The current study examined the utility of EEG based cognitive metrics for tracking workload and engagement in differentially experienced ATC students over the course of a dynamic, five phase high-fidelity TRACON simulation. Performance results indicated that both low and high experienced students performed similarly on the scenario, with the most errors occurring during Phase 4. Phase 3 had the most aircraft released along with uncontrolled aircraft, making Phase 3 more likely to induce errors. However, the most errors occurred during Phase 4. This was likely the result of a progressive buildup of traffic from the high workload of Phase 3, causing more errors further along the line in the scenario. The B-Alert workload metric demonstrated a trend that mirrored the predicted workload of the scenario. That is, the metric increased from Phases 1 through 3 and then decreased after Phase 3. Thus, this metric was able to track changes in cognitive workload throughout the simulation. This study provides further evidence for potential dissociations that can occur with performance and other measures of workload. The most errors occurred during Phase 4, however, peak EEG workload was observed in Phase 3. This finding underscores the importance of assessing operator workload with different measures to acquire a holistic picture of the operator’s state. The B-Alert engagement metric did not significantly fluctuate throughout the scenario for either experience group, indicating that controllers remained consistently engaged throughout the scenario.

Although statistical analysis did not demonstrate experience main effects or phase by experience interactions with respect to the EEG workload and engagement metrics, visual analysis of the data shows a trend toward lower experience students exhibiting higher workload during Phases 1 through 3, as well as higher engagement throughout the scenario. The slightly increased workload observed in lower experienced students likely reflects less of an ability to mobilize cognitive resources efficiently (Berka et al., 2007). Moreover, the elevated engagement in lower experienced students is likely the antecedent to increased workload. Low experience controllers must maintain a higher degree of focused attention to maintain performance and allocate attentional resources, a cost not revealed by simulator performance metrics. The trend in engagement is similar to previous research demonstrating higher EEG indices of engagement in novices compared to experts during task performance (Johnson et al., 2014). Future studies should utilize stronger manipulations of controller experience (e.g., professional controllers vs. students). In using upper and lower leveled students, the skill level difference between the two groups may not have been strong enough to produce statistically significant differences between groups.

References

Berka, C., Levendowski, D. J., Lumicao, M. N., Yau, A., Davis, G., Olmstead, R. E., … Craven, P. L. (2007). EEG correlates of task engagement and mental workload in vigilance, learning, and memory. *Aviation, Space, and Environmental Medicine*, *78*(5), B231- B243

Desmond, P. A., & Hoyes, T. W. (1996). Workload variation, intrinsic risk and utility in a simulated air traffic control task: Evidence for compensatory effects. *Safety Science*, *22*(1-3), 87-101.

Field, A. (2009). *Discovering statistics using SPSS.* Thousand Oaks, CA: SAGE Publications Inc.

Hancock, P. A., & Warm, J. S. (1989). A dynamic model of stress and sustained attention. *Human Factors, 31*(5), 519-537.

Johnson, R. R., Stone, B. T., Miranda, C. M., Vila, B., James, L., James, S. M., … Berka, C. (2014). Identifying psychophysiological indices of expert vs. novice performance in deadly force judgment and decision making. *Frontiers in Human Neuroscience*, *8*, 1-13. doi: 10.3389/fnhum.2014.00512

Loft, S., Sanderson, P., Neal, A., & Mooij, M. (2007). Modeling and predicting mental workload in en route air traffic control: Critical review and broader implications. *Human Factors*, *49*(3), 376-399. doi: 10.1518/001872007X197017

Parasuraman, R. (2015). Neuroergonomic perspectives on human systems integration: Mental workload, vigilance, adaptive automation, and training. In D. A. Boehm-Davis, F. T. Durso, J. D. Lee. (Eds.), *APA handbook of human systems integration* (pp. 163-176). Washington, DC: American Psychological Association.

Saxby, D. J., Matthews, G., Warm, J. S., Hitchcock, E. M., & Neubauer, C. (2013). Active and passive fatigue in simulated driving: discriminating styles of workload regulation and their safety impacts. *Journal of Experimental Psychology: Applied*, *19*(4), 287-300.

Signal, T. L., Gander, P. H., Anderson, H., Brash, S. (2009). Scheduled napping as a countermeasure to sleepiness in air traffic controllers. *Journal of Sleep Research*, *18*, 11-19. doi: 10.1111/j.1365-2869.2008.00702.x

Vogt, J., Hagemann, T., & Kastner, M. (2006). The impact of workload on heart rate and blood pressure in en-route and tower air traffic control. *Journal of Psychophysiology*, *20*(4), 297-314.

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

Wickens, C., & Tsang, P. (2015). Workload. In D. A. Boehm-Davis, F. T. Durso, & J. D. Lee (Eds.), *APA handbook of human systems integration* (pp. 277-292). Washington, DC: American Psychological Association.