INTERNATIONAL JOURNAL OF HUMAN–COMPUTER INTERACTION
https://doi.org/10.1080/10447318.2023.2208991

Biometrically Measured Affect for Screen-Based Drone Pilot Skill Acquisition

Fatemeh Dalilian and David Nembhard

University of Iowa, Iowa City, IA, USA

ABSTRACT

Drones have been used increasingly to aid in Industry 4.0 activities including inspection of the nation's infrastructure. We investigate several potential underlying affective behaviors related to drone pilot skill acquisition, with the eventual goal of developing methods to enhance human performance. We employ Electroencephalography (EEG) and Eye tracking instrumentation to measure human affect in a series of simulated drone piloting experiments to examine performance using behavioral variables, controller input variables, as well as measures of individual cognitive ability. Current results show that task difficulty impacts the performance/learning process and varies by the nature of the task. The behavioral and biometric measures associated with performance/learning varied significantly among activities. We conclude that drone specifications and training requirements can and should be calibrated to the drone mission. In addition to developing specifications and training requirements, psychological and behavioral measures can also serve as theoretical foundations for modeling complex tasks.

KEYWORDS

Drone pilot training; human-drone interaction; flight simulation

1. Introduction

Rapid advances in technology have led to ongoing industrial transformation, often referred to as the Fourth Industrial Revolution or Industry 4.0 (Dalenogare et al., 2018). In this domain, new technologies are being introduced and adapted at an accelerating pace to transform the future of work. A world economic forum report estimated that a third of job requirements in 2025 will involve technology-related skills not yet considered crucial for today's jobs (Schwab & Zahidi, 2020). Retraining 50% of existing employees will be necessary by 2025 to keep up with the pace of technological advancement (Schwab & Zahidi, 2020). Thus, a strong commitment to workforce reskilling/upskilling is required by both individuals and organizations to prepare for what the future of work will entail (Li, 2022). In industry 4.0, drones, alternately unmanned aerial vehicles (UAVs), are one of the technologies transforming agriculture, construction, inspection, logistics, and many other industries due to their operational efficiency and economic benefits (Javaid et al., 2022). The drone industry market size is estimated to grow from 4.4 billion in 2018 to 63.6 billion by 2025 (Intelligence, 2021), creating 100,000 jobs in the United States (*Economic Report*, 2017). However, the training requirements for new skills required for this and other future work have not been extensively researched (Kopp et al., 2021).

The use of drones as an Industry 4.0 job enabler generally reduces physical demands but introduces new sources of cognitive demand (Kong, 2019; Weiss et al., 2021). In scenarios where the same individual performs the inspection and operates the drone, an increased cognitive demand can be attributed to multitasking—with piloting as the primary task and performing the task(s) previously done manually as a secondary task (Razafimahazo et al., 2021). In scenarios where technologies, pilots, and the others collaborate, the increased cognitive demand may be due to shared or shifted cognitive and attentional resources among multiple tasks required for cooperation (see Cole et al., 2010; Crandall et al., 2005; Kruijff et al., 2012; Roldán-Gómez et al., 2021). In this multitasking paradigm, performance is sensitive to cognitive load and attentional resource availability (Brown & Bennett, 2002). Achieving "good" results on the primary task with minimal cognitive effort enhances system performance (Poldrack et al., 2005). Thus, effective pilot training is critical for safe drone deployment and efficient performance in practice (Kadir et al., 2019; Pacaux-Lemoine & Trentesaux, 2019).

Given the complex nature of skill acquisition for fine motor control tasks, drone piloting competency should not only take performance metrics into account but also cognitive, behavioral, and physical measures. Thus, a multimodal approach is a natural direction for assessing trainees' learning status (Mu et al., 2020). Correspondingly, data collected from multiple sensors and instruments have the potential to provide insight into aspects of drone piloting performance and allow for a deeper understanding of how drone skills are developed.

In this study, a drone pilot learning process was examined by identifying correlates of specific cognitive states and behaviors related to drone control performance. The combination of biometric signals, behavioral measures, difficulty levels, and drone tasks enriches the understanding of

CONTACT Fatemeh Dalilian fatemeh-dalilian@uiowa.edu University of Iowa, Iowa City, IA, USA

dynamic interaction between pilot state, mission, and drone control. Learning-related constructs can inform training designs (real-time feedback, training hours, etc.) and performance evaluation tools (Wong et al., 2014).

1.1. Skill acquisition

Many motor-skill learning models suggest that "practice makes perfect" or at least that "practice makes progress." Through practice, certain subcomponents of a skill become automatic, and task execution becomes relatively effortless (Jaquess et al., 2018). During the early stages of learning, performance accuracy and stability improve rapidly, and then progress more gradually. During this process, task execution advances from an attention-demanding early phase to an automatic phase (Fitts & Posner, 1967). When tasks are performed automatically without the need for executive control, known as *automaticity*, the learned task is performed with minimal cognitive load. Therein, task performance is not hindered or attenuated by a secondary task (Posner, 1975). Figure 1 details the skill acquisition stages and their corresponding input and output (Furley & Memmert, 2010).

Cognitive load depends on limited individual attentional resources, as commonly measured based on working memory capacity, a construct of individual cognitive ability (Baddeley & Hitch, 1974). Conway et al. (2002) examined the interrelationships among cognitive ability constructs such as working memory capacity, processing speed, short-term memory capacity, spatial ability, and selected working memory capacity to be the best predictor of general cognitive ability in young adults (Shipstead et al., 2012). In complex tasks such as surgical operations, aircraft piloting, operating manual space tasks, driving, etc., it has been consistently found that performance correlates with individual cognitive abilities (Du et al., 2015; Rode et al., 2014; Rohde & Thompson, 2007). Higher working memory capacity tends to be associated with better motor learning rates both in adults (Bo et al., 2009) and children (Buszard et al., 2017).

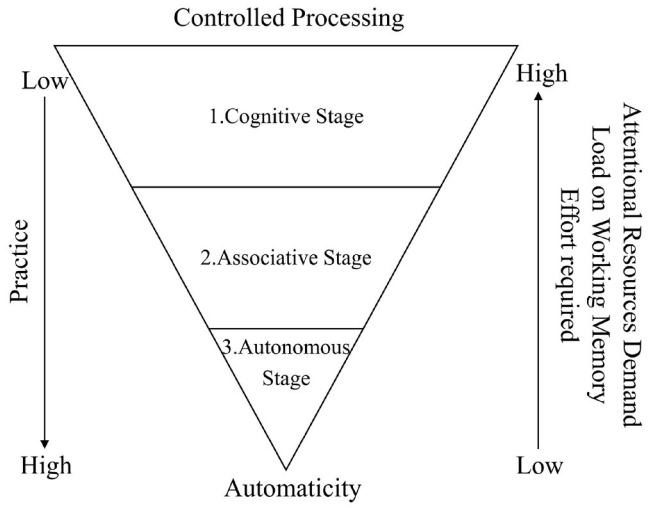

Figure 1. Fitts and Posner (1967) skill acquisition model (adapted from Furley & Memmert, 2010).

During the autonomous stage of learning (Figure 1—Stage 3), cognitive processing requirements are limited, and neither higher cognitive ability nor greater working memory capacity can guarantee greater performance. However, research has also shown that cognitive ability cannot be fully compensated for by practice, particularly for complex tasks such as air traffic control and laparoscopic surgery (Du et al., 2015). It has been suggested that this is due to higher working memory capacity and individuals being less responsive to irrelevant information, and having better attentional control, thereby helping them to achieve better results (Engle, 2002). Gevins and Smith (2000) demonstrate superior sustained focus among those with high cognitive ability based on EEG biometric measures. In their study, frontal midline theta band power was higher in the high cognitive ability group when they were performing working memory intensive tasks. The EEG alpha band power in parietal regions was also higher suggesting that they make greater use of parietal regions, more commonly associated with autonomous control, rather than relying solely on frontal regions, associated with attentional control.

1.2. Biometrics and behavioral measures

The employment of subjective measures and self-report measures is vast and generally useful for research. However, their use can limit the efficacy of measurements of behavior during real-time tasks (Schmidt et al., 2009). Querying participants during tasks creates a secondary task load likely to affect primary task performance. Further, humans often make incorrect judgments regarding their own mental state (Sano et al., 2018). Objective measurement, where possible, such as the use of biometrics and physiological measures generally involves manifestations that cannot be controlled voluntarily (Lohani et al., 2019). In the science of learning, biometric and behavioral measures have been extensively used to identify the associative changes in the brain and behavior that occur with learning (Nakano et al., 2013; Willingham, 1999). Biometric and behavioral measures can be collected using various sensors and instruments. For example, Inertial Measure Units (IMU) use accelerometers and gyroscopes to measure linear movements as well as the rotation of body parts (Fong et al., 2008). Surface electromyography (EMG), which records electrical activity in muscles, is used to evaluate muscle strain and fatigue during operations and to respond to environmental stimuli (Farina et al., 2016). By measuring heart rate and heart rate variability, electrocardiograms (ECGs) provide information regarding stress levels (Goodie et al., 2000). Using Thermal Imaging, it is possible to capture changes in skin temperature on the face which is linked to cognitive load and task difficulty (Abdelrahman et al., 2017). Eye movement tracking (EMT) is effective at measuring a range of behaviors including the level of attention and effort using pupillary dilation as well as areas of attention and search patterns (De Cecilio de Carlos, 2019). The electroencephalogram (EEG), which measures electrical activity associated with various regions of the brain, is used to quantify and assess focus,

cognitive load, and emotional responses (Hajcak et al., 2010). By integrating any number of these biometric and behavioral sensors, richer insight into individuals' cognitive states can be gained (Liu & Wickens, 1994; Merceron et al., 2016). Several drone pilot training studies have successfully implemented a multimodal approach toward this end. For example, Corichi et al. (2017) modeled drone pilot altered states in real time using skin conduction, breathing, and heart signals. Pham et al. (2020) measured drone pilots' engagement levels using the pilots' EEG metrics and provided this measurement as feedback. They found that the quantified score of pilots' engagement level invoked self-regulation that motivated them to improve focus. Having multivariate affective measures can provide the ability to address a far wider range of issues about varied drone flight regimes and tasks.

Given that the coordination of brain, eyes, and hands is vital for drone piloting (Pfeiffer & Scaramuzza, 2021), three data collection systems were chosen in this study. Each of these represents a common standard in their respective domain. EEG was used to measure brain electrical signals, eye tracking was used to record eye movement behavior along with the involuntary pupillary response, and controller inputs were used for hand movements and muscle activity. Other common sensors and measurement approaches such as EMG and IMU were considered but not employed in the current study. Since uncontrolled changes in muscle activity and hand position can be confounded with physiological variations while using the drone controller, EMG would likely have limited utility (Lohani et al., 2019). A fundamental limitation of IMU is the drifting problem, where the estimated position accumulates tracking error over time (Polfreman, 2018). Further, fine-grained hand pose tracking requires multiple IMUs on a pair of gloves (Chen et al., 2021) which is somewhat invasive and potentially biasing, during drone flight. Control input data collection is transparent to the participant, and a low-noise approach for measuring hand movements and control behavior.

1.2.1. Electroencephalography

Human learning involves the modification of synapses among neurons and the formation of new memories (Rioult-Pedotti et al., 2000). Correspondingly, EEG is potentially useful for evaluating information processing as a function of activity. Changes in EEG signals reflect a substrate of psychological variables such as attention, cognitive load, and arousal. Thus, metrics can be used to characterize humans' mental states and correspondingly, performance (Besserve et al., 2008). Nakano et al. (2013) showed that not only practice and skill acquisition is associated with EEG measures, but also greater changes in EEG measures (alpha and beta powers in the parietal lobe locations), which indicates greater improvement in motor skill performance.

Many studies have used EEG frequency band powers related to specific functional characteristics to predict task performance, including finger tapping, video games, surgery, driving, and piloting (Astolfi et al., 2005; Plazak et al., 2019; Sheikholeslami et al., 2007). In general, standard frequency bands and their approximate spectral boundaries have been used: delta (1–3 Hz), theta (4–7 Hz) and alpha (8–12 Hz), beta (13–30 Hz), and gamma (30–100 Hz). Elevated alpha power during a cognitive-motor task practice has been associated with a decrease in task-related cognitive load (Kerick et al., 2004; Zhu et al., 2010). Attention to task-appropriate cues or focused attention has been associated with high theta power in the frontal midline region (Smith et al., 1999). Beta power has been linked to central nervous system activation, alertness, arousal, vigilance, and visual attention (Andreassi, 2010; Ramautar et al., 2013). Kamiński et al. (2012) found that higher activations in the beta band correlate with increased alertness, and are related to faster responses to target visual stimuli. During lower vigilance, alpha rhythm tends to diminish (alpha drop-out) and beta power increases—a state similar to that measured by EEG during intense mental activity during an open-eyes session (De Gennaro et al., 2001; Tanaka et al., 1996, 1997).

EEG measures have been considered for drone pilot training as well. In a drone simulator training study, participants' mental workload was determined using EEG measures with an average accuracy of 87% (Gu et al., 2022). By using drone-pilot EEG signals, Park et al. (2021) developed a system to detect and alert operators if they begin to become inattentive. In another study, alpha and beta wave metrics were used to classify the emotional states of drone pilots as quiet or very tense states (Olivares-Figueroa et al., 2021). Jao et al. (2021) also showed that EEG signals are reliable features for discriminating between cognitive states that covary with drone task difficulty.

1.2.2. Pupillometry

Eye-tracking technologies have been effectively used for studying complex task learning (Fletcher et al., 2017; Krejtz et al., 2018). Several measurements of visual behavior including dwells, blink durations, blink rates, pupil diameters, and eye closure times have been shown to be useful in aviation research (Peißl et al., 2018). Among eye tracking metrics, pupillary response was selected to reflect information processing demands in beyond direct physical metrics (Backs & Walrath, 1992). Pupillometry has gained considerable attention as an objective measure of affect related to cognitive features, underlying skill learning, dynamic decision-making, expert *vs.* novice behavior, multitasking, and complex task performance (Peysakhovich et al., 2015; Qiao et al., 2022; Tichon et al., 2014). A review study by van der Wel and van Steenbergen (2018) supported that pupil dilation reflects effort exerted in response to increased cognitive demand during cognitive control tasks. Thus, greater task demands are associated with greater pupil dilation and once task demands exceed individual cognitive capacity, pupil dilation tends not to increase further.

The relationship between pupil dilation, task difficulty, and performance has been studied in several flight simulation studies, with at-times conflicting conclusions. Sibley et al. (2010) using both EEG and eye-tracking metrics concluded that pupil dilation was a good indicator of drone pilot skill acquisition. In a subsequent study, Sibley et al.

(2011) showed that pupil diameter significantly decreased after each block of simulated drone task practice, and then increased with greater task difficulty. There was no consistent link between pupil dilation and difficulty levels in flight simulations in a later study by Tichon et al. (2014). For participants who reported elevated levels of anxiety under high difficulty flying conditions, pupil size increased, yet decreased for participants whose anxiety levels did not increase when moving from easy to more complex flying conditions (Tichon et al., 2014). From this, it appears that the effect of task difficulty (higher workload) on pupil dilation may be influenced by individual cognitive processes and the specific task at hand.

1.2.3. Controller inputs

Mathematical modeling of automobile driver steering behavior has led to interesting findings about driver performance during lane keeping (Hess & Modjtahedzadeh, 1990), following target path curvature (Donges, 1978), emergency response (Comolli et al., 2020), collision avoidance (Markkula et al., 2014), and lane change maneuvers (van Winsum et al., 1999). Studies have for decades divided the steering task along two dimensions: *guidance* and *stabilization* (Donges, 1978; Land, 1998; Salvucci & Gray, 2004; Zhao et al., 2020). Guidance input involves anticipatory open-loop control executing a steering trajectory. Stabilization is a closed-loop control input to correct for deviations with respect to a target route (Donges, 1978). A control-theoretic model for steering behavior was also introduced by Hess and Modjtahedzadeh (1990), who included low-frequency and high-frequency compensation components. Drones are commonly guided, controlled, and navigated via a ground control station that often includes a tablet-sized screen, and two joystick controls, one for rotating the drone and one for propelling it (Kang et al., 2018). Within this common general setup there are countless implementations. Further, there are many designs outside of this, including collaborative designs for teams, and those with a single side stick control. Controller input behavioral patterns are potentially useful for measuring the performance of pilot-drone systems in a similar manner to those of other operator-controlled systems, including automobile driver-vehicle systems more broadly.

Similar dual-input systems have been researched in the context of screen-based games. Li et al. (2016) had participants control a target toward the center of the screen using a joystick to determine control response amplitudes. Across input perturbation frequencies, action gamers had a larger mean control response amplitude than non-action gamers (Li et al., 2016). In a related study, a joystick was used to control a horizontal luminance-defined line whose position was randomly perturbed (Li et al., 2005). Frequency analyses were performed to compute both the closed-loop and open-loop functions for the participants' control response. They found that a higher contrast screen reduced phase lag and increased open-loop gain. Peng et al. (2010) examined how visual path information affects heading control concluding that when path information is available, low-frequency (0.3 Hz) drift is reduced. In a study by Rizzi et al. (2021), a preview with a window of 0.3–0.6 s was compared with a preview with a full 1-s preview. Therein, a Fourier analysis of participants' joystick movements was used to represent responses to sinusoidal roadway perturbations, an indication of attention allocated to different roadway preview locations. More specifically, participants showed greater joystick movement when the window was 0.3 s into the future, suggesting they tracked the higher frequency components of the roadway more accurately. A total system error has been studied by Congress et al. (2018). Maza et al. (2010) provided a comparative review of various enabling technologies that are informative and supportive of commonly used input and measurement systems in the literature, as well as those used in the current study.

1.3. Research questions

Schmidt et al. (2022) conducted a study providing insights into the skills and training needs of civil drone pilots and suggested a range of technical, operational, and theoretical competencies. In terms of training, they noted that essential competencies should be over-trained to ensure mastery. Thus, there is a need for initial and frequent recurrent training that is tailored to the specific needs and skills of individual pilots. To improve drone pilot training, various interventions have been proposed. Gamification is often effective for training and aids the development of scenarios and strategies for creating virtual environments tailored to specific missions. For example, some have simulated the use of drones for agriculture, search and rescue, and increasing motivation (Cardona-Reyes et al., 2021). A virtual reality (VR) environment was designed by Sakib et al. (2021) to closely mimic a practical setting with regard to spatial perception and dimensions, as well as drone features for training purposes. They examined physiological characteristics of participants, including stress, and heart rate, which indicated no significant differences between simulation and the real-world environment. This suggests that participants had similar experiences in both settings. Similarly, Albeaino et al. (2022) found that a simulation-based environment and a real-world environment for training drone-building inspection were comparable in terms of in-flight workload demand as measured by NASA-Task Load Index (TLX).

Inspector-pilot task performance was studied to assess system usability (Eiris et al., 2021), with results showing that participants could identify prior pilot inspection strategies based on flight paths, drone orientations, and inspection marker affordances. Moreover, trainees were able to quickly identify difficult-to-maneuver areas. Ribeiro et al. (2021) implemented a real-time Augmented Reality (AR) experience prototype for drone pilot training as a cost-effective solution for pilot training and mitigation of crash-related technophobia.

Table 1 summarizes four research questions examined in this study alongside related hypotheses. Within the extant literature, biometric and behavioral measures and findings related to drone piloting skill acquisition differ, despite the

Table 1. Research questions and hypotheses.

	Research question	Hypotheses
1	Are biometric and behavioral measures associated with drone piloting skill acquisition task-specific?	The biometric and behavioral factors significantly affecting completion times for vertical control task and transfer and land task are different.
2	Does cognitive ability moderate the effect of biometric measures on performance?	There is a non-zero interaction coefficient between biometric measures and cognitive ability scores on completion times.
3	Does the amount of control input behavior impact performance?	One or more input behaviors have a coefficient that is not zero in the linear equation with completion time as the response variable.
4	How does increased payload (weight) affect task performance?	There is a significant difference in the mean completion time before and after the payload increase.

general potential of these measures to facilitate learning improvements. An analysis by Senoussi et al. (2017) showed that high alpha-band power in frontal lobe electrodes was positively related to poorer pilot performance as measured by slower reaction times. However, in a 2-hr drone pilot training session study, the first 40 min, alpha power increased significantly and then dropped significantly afterward (Roy et al., 2016). The beta-band power increased significantly during the last 20 min. Gu et al. (2022) showed that measuring mental workload during drone piloting training involves the consideration of cognitive states at the time. Thus, a single EEG power spectral density indicator may not be sufficient to accurately describe a person's mental workload at all learning stages. Gu et al. showed that before the development of mental schema, only theta power increased with increased task difficulty. During mental schema formation, task difficulty resulted in a reduction of frontal theta power as well as parietal alpha power and an increase in central beta power. At a more mature stage of mental schema, there was still a significant difference between the frontal theta power and central beta power at different difficulty levels, but the parietal alpha PSD was unaffected. They attributed the differences in EEG workload indicators to mental schema progression (Gu et al., 2022). We hypothesize that differences in biometric and behavioral indicators of drone pilot performance may be due to considerable differences in the drone missions and the human behaviors that play a mediating role in performance. In the current study, participants were trained for two different drone tasks to address research question *Q1*: *Are biometric and behavioral measures associated with drone piloting skill acquisition task-specific?* The effects of cognitive load, attention, vigilance, and effort on pilot trainee performance will be examined for each task separately.

Drone piloting is a largely cognitive task. Thus, it is likely that individual working memory capacity should have an effect on how biometric and behavioral measures relate to individual performance. For example, Ahern and Beatty (1979) showed that intelligence moderates the relationship between performance and pupil dilation. Individuals with higher levels of intelligence exhibited smaller pupil dilation and responded more accurately to all levels of task difficulty (Ahern & Beatty, 1979). They concluded that intelligent individuals process information more efficiently, therefore requiring less effort. Correspondingly in the drone piloting domain, we will assess *Q2: Does cognitive ability moderate the effect of biometric measures on performance?*

In addition to biometrically measured affect, quantitative measures of joystick movement behavior are potentially useful for representing and identifying control behaviors associated with performance levels. In the current study, a standard two joystick controller setup is used to assess *Q3: Does the amount of control input behavior impact performance?*

Many of the most basic drones are equipped with multiple sensors that allow for multiple roles and missions. Depending on the payload, the aerodynamics and controllability of the drone will change, leading to the research question, *Q4: How does increased payload (weight) affect task performance?*

2. Methodology

A series of controlled experiments using drone flight simulations in a laboratory setting were conducted. Individuals' brain electrical activity, eye movements, and control behavior during the training sessions for two different drone tasks were recorded to explore how biometrically assessed affect and behaviors influence performance outcomes and whether task difficulty moderated those effects. EEG power within specific frequencies was estimated to determine individuals' states including cognitive load, focused attention, and vigilance. Levels of effort were assessed by recording eye events using pupillary dilation. Control behavior was evaluated by the controller inputs signal, and working memory was assessed using an OSPAN (Operation Span) evaluation score. Pilot trainees' operational readiness was investigated considering their working memory capacity, and their affective and behavioral characteristics as well as task type and task difficulty. Figure 2 summarizes the set of data collected, along with biometrics and behavioral metrics, and subsequent evaluative variables. In Figure 3, a participant is shown during the experiment.

Accounting for practical factors such as the number of participants and experiment time, a total of 12 biometrics and behavioral measures relevant to the research questions were selected (Table 2) and detailed in the sections that follow. A stepwise mixed model regression was used with participants as a random effect, behavioral and affective metrics as independent variables, and performance as the dependent variable.

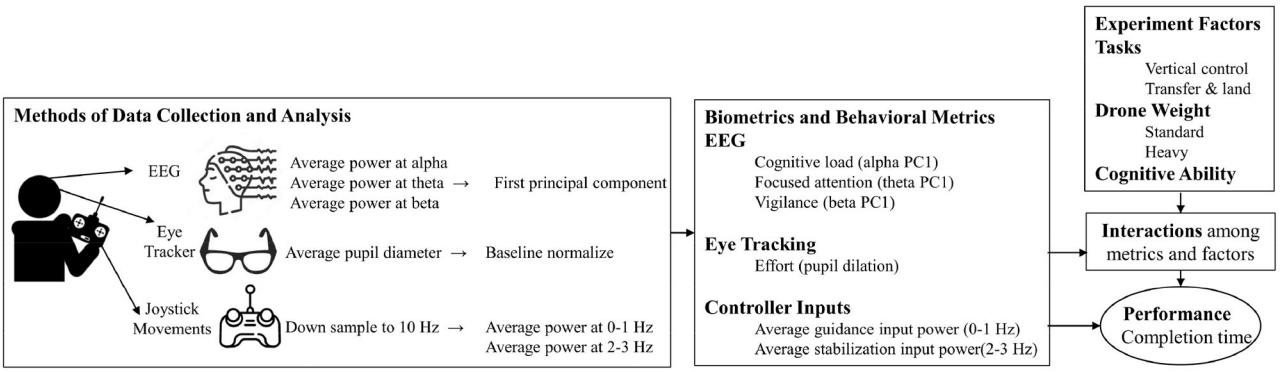

Figure 2. Experimental data collection and measures.

Figure 3. A participant during the experiment.

Table 2. Biometric and behavioral measures and associated psychological features.

Biometric and behavioral metrics	Psychological features
Alpha PC1	Cognitive load
Theta PC1	Focused attention/enhanced attention control
Beta PC1	Vigilance/arousal/alertness
Pupil dilation	Effort
OSPAN score	Working memory capacity
Throttle power 0–1 Hz	Vertical guidance input size
Throttle power 2–3 Hz	Vertical stabilization input size
Aileron power 0–1 Hz	Sideward guidance input size
Aileron power at 2–3 Hz	Sideward stabilization input size
Elevator power at 0–1 Hz	Forward/backward guidance input size
Elevator power at 2–3 Hz	Forward/backward stabilization input size
Rudder power at 0–1 Hz	Turn guidance input size
Rudder power at 2–3 Hz	Turn stabilization input size

2.1. Participants and study design

Roughly 90% of drone pilots recommend that novices practice drone piloting by initially using simulators for the early learning phases since crashes are highly probable for drone pilot trainees (Tezza et al., 2021). *Aerosim RC,* a commercial pilot training simulator, was employed as the experimental platform reducing risks and limiting sources of noise. Using institutional review board protocols, fifteen graduate and undergraduate engineering students at the University of Iowa, participated in a series of *repeated measures experiments* as a variance reduction technique (e.g., Girden, 1992). Novice students were selected to best understand the early phase learning behaviors among a common cohort. Thus, none of the participants had prior drone piloting experience, they all tended to have less than expert levels of experience with video games, but none were familiar with the flight dynamics of drones. The average age of the participants was 21.3 years (std dev. = 5.08), and 10 were males and 5 were females.

Participants watched a short training video on the two piloting tasks: "*vertical control*" and "*transfer and land*." Since low cognitive load during training enables individuals to improve their performance more rapidly after transfer to a high workload (Gonzalez, 2005), participants began with the less complex, lighter—standard drone. After finishing 8 repeated trials with the standard drone successfully, the piloting task difficulty was increased by increasing the drone weight by 10% and subsequently performing eight additional trials. Drone flight dynamics are well-known to be more difficult to manage with even modest increases in drone weight, other parameters remaining equal. If a crash occurred at any point, the trial was repeated. The Experimental steps are summarized in Figure 4.

The experiment was conducted inside a sound-proof booth. The distance from participants' eyes to the screen was ∼75 cm, with stimuli on a 24-in monitor, and a corresponding angular size of 54.7 degrees. Learning performance for each task was measured via session completion time.

2.2. Experimental measures

The experimental measures as summarized in Figure 2, are further detailed alongside their corresponding behavioral indications in the subsections that follow.

2.2.1. Working memory capacity

Before the drone piloting tasks, participants' working memory capacity was measured using an automated OSPAN test. Unsworth et al. (2005) validated this version of OSPAN showing its internal consistency, test-retest reliability, and a strong correlation with other measures of working memory capacity.

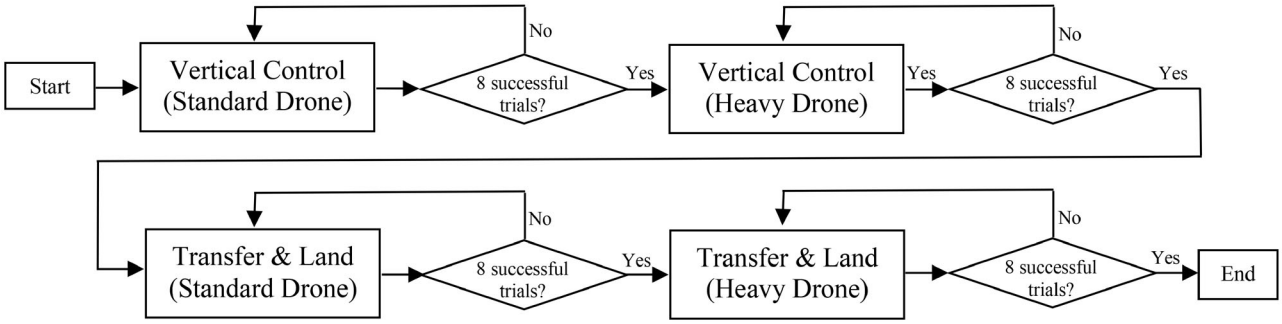

Figure 4. Experimental procedure.

2.2.2. EEG band powers

The B-alert X10 EEG headset, with 256 Hz sampling frequency was used to collect participants' brain electrical activity. Nine electrode locations were measured, using the standard international 10–20 system (*F3*, *Fz*, *F4*, *C3*, *Cz*, *C4*, *P3*, *Poz*, *P4*) referenced to the earlobes. The EEG signals recorded from all nine electrodes were internally decontaminated using ABM decontamination, an algorithm that removed EMG, EOG, excursions, saturations, and spike artifacts. Saturations are defined as five consecutive data points with amplitude changes within the amplifier saturation range. Excursions were characterized by sudden or constant changes in EEG amplitudes across three, four, or five data points with only one zero crossing in a 128-point region. Data from saturated and excursion periods were replaced with zero values. Spikes were identified as bidirectional amplitude fluctuations over 40 microvolts, measured across three, five, or seven data points, and their values were interpolated in the decontamination process. The EEG signal was deconstructed using wavelet transforms into 6 wavelet bins (i.e., 0–2, 2–4, 4–8, 8–16, 16–32, 32–64, and 64–128 Hz). Subsequently, a stepwise regression was used to select the wavelet bins for the discriminant function classifying data points contaminated with eye blinks. Wavelet coefficients in the eye blink range in the 0–2, 2–4, and 4–8 Hz bins were replaced with artifact-free mean wavelet values in each channel of EEG. The EEG signal is recomposed using the wavelet bins from 2 to 64 Hz as suggested by Berka et al. (2005).

For each participant, all key events associated with the tasks were annotated, such that the training sessions' start and finish times were marked and registered. MatLab (2020) was used for signal processing, including EEG Power Spectral Density, which was calculated during each training session using Welch's method with a 2-s Hamming window and 1-s overlap. The average power was approximated by integrating the power spectral density (PSD) estimate contained in the beta, alpha, and theta frequency ranges. Similarly, the beta, alpha, and theta average powers were computed during the last 60 s of a 2-min closed-eyes rest period as a baseline, followed by a decibel normalization of the average power relative to the baseline. This was performed for each of the nine electrode locations and each of the alpha, theta, and beta frequencies.

For each of the frequency band powers, a Principal Component Analysis (PCA) was applied. Since PCAs reduce the dimensionality of the multivariate data through rotation to a new basis using eigenvectors, the new high-variance principal axes form a potentially more efficient set of PCA variables for use in subsequent analyses. This has also been shown to be effective in removing oculomotor artifacts in EEG (Kaczorowska et al., 2017). In this study, the first components of alpha, theta, and beta decibel powers tended to explain 91.5, 75.8, and 79.2% of the variation, respectively. Alpha power loadings on the first principal component for alpha powers (alpha PC1) were all positive values between 0.316 and 0.344. Greater alpha power was associated with less cognitive load (De Gennaro et al., 2001; Tanaka et al., 1996, 1997). To simplify interpretations, the alpha power signs were reversed such that greater alpha PC1 would indicate greater cognitive load. Beta PC1 (The first principal component for beta powers) loadings were between 0.313 to 0.353. Beta power is an indication of alertness, vigilance, and visual perception (Andreassi, 2010; Kamiński et al., 2012; Ramautar et al., 2013). Electrodes' theta power loadings for theta PC1 were all positive values between 0.271 and 0.366 with heavier central and parietal loadings. Increased theta power is interpreted as more attention, higher task demand, and more task difficulty (Ishii et al., 2014; Smith et al., 1999; Travis & Shear, 2010).

2.2.3. Pupil dilation

Tobii Glasses-3 eye-trackers were used to collect participants' eye movement metrics at a 120 Hz sampling rate. A calibration procedure was performed to maximize data quality, wherein participants look at a target card held at arm's length (0.5–1 m). The calibration procedure optimizes the performance of the gaze estimation algorithm of the eye-tracker (i.e., the 3D eye model) (TobiiConnect, 2022). All participants' calibration was completed successfully before the experiments, and the eye tracker recorded the entire experimental process.

Pupil dilation is an indication of cognitive load and the exerted effort to complete a task. Pupillary diameter recorded from both eyes during the OSPAN was averaged as the baseline pupil diameter. The baseline pupil diameter for each participant was subtracted from their average pupil

diameter during each session to find the pupil diameter change during each session.

2.2.4. Controller input behavior

There remains considerable room for development before autonomous drone flight operation and decision-making surpasses or matches the judgment capacity of experienced inspectors/engineers (Zhang et al., 2022). The reliability and liability risks of sensors and artificial intelligence are major concerns for automated drone missions. However, sub-processes can be successful at various levels of autonomy (Zhang et al., 2022). Currently, many drone applications, such as structural inspections, require manual operation since autonomous flight limits the ability of human operators to intervene in real-time if issues arise during flight (Manka, 2022).

Dual joystick remote controllers are a common design for manual drone control (Kang et al., 2018). Consumer drones with on-screen virtual controllers tend to be lighter, smaller, and comparatively lower-priced, but they are also more difficult to navigate (Kang et al., 2018). Thumb-operated joysticks are generally considered to be more user-friendly as they offer a more intuitive and physical control interface (Rupp et al., 2013). The current experiment was conducted using a hand-held remote control with two joysticks. The left joystick was for vertical motion (throttle) and changing the drone head position (Yaw). The right stick was for forward/backward (pitch) and sideway motion (roll).

Stick positions (throttle, yaw, pitch, and roll) were recorded with a sampling rate of 40 Hz and then were down-sampled to 10 Hz. Using a two-level input, namely guidance, and stabilization, it is possible to capture pilots' behaviors, performance, skill level, and possibly recommend training requirements. For example, if a trainee's control behavior shows deviation on one task but not on others, a recommender system could suggest additional task-specific practice. The average input power imparted to the joystick in the frequency ranges of 0–1 and 2–3 Hz was determined as an indication of guidance level and a stabilization level input size. The notion of guidance and stabilization level inputs was inspired by driver steering behavior literature (Donges, 1978; Lehtonen et al., 2014; Mole et al., 2016).

In the drone piloting literature, the concept of guidance level and stabilization level input has not been widely studied. *Guidance* is defined as inputs with frequencies lower than 1 Hz and *stabilization* inputs have frequencies over 2 Hz. A control input between 1 and 2 Hz was defined as an occasional input, regarded as an open-loop anticipatory control, and served as a bound for 0–1 Hz for guidance inputs. Higher frequency inputs (>2 Hz) were considered to be closed-loop control inputs that compensated for deviations from an intended target. Two to 3 Hz input frequencies were assumed to be related to stabilization level inputs. Additional research in other domains beyond automobiles and drones may confirm these general guidance and stabilization control behavioral inputs.

2.3. Analysis

This study employed repeated measures of 15 participants leading to two categories of variation, *between-subject*, and *within-subject*, which were separated using mixed models. Within-subject measures (e.g., drone-task completion times at each session as in Figure 2) are not likely to be independent, rather they are correlated. The participants were regarded as independent random samples from the population. Thus, mixed-model regression was used to analyze and distinguish between individual-level and group-level variance (Van Dongen et al., 2004). To find the important behavioral and biometric measures associated with learning as well as the moderating effect of cognitive ability (Q1–Q3), a mixed model regression was used with participants modeled as random effects, behavioral and affective measures as independent variables, and performance as the dependent variable. Biometric and performance measures tend to have relatively high variability among individuals (Basile et al., 2007). To prevent falsely generalizing between-subject differences to within-subject effects, the independent variables were normalized (see e.g., Van de Pol & Wright, 2009; Wong et al., 2014). Individual differences in the dependent performance variable were modeled using a mixed model regression, and a bi-directional stepwise procedure was used given the multiple independent variables under consideration (Kramer, 2004).

As completion times for different sessions and complexity levels are observed within subjects, the samples were analyzed using paired-variable statistics. To examine how practice affects task completion times and how increasing difficulty levels affect performance (Q4), three pairwise paired *t*-tests were performed for each task: one for the first and last sessions of standard drone training, one for the last session of standard drone training and the first session of heavy drone training, and one for the first and last sessions of heavy drone training. Before the tests, outliers were identified, and the normality of the data was tested using the Shapiro-Wilk test. For the difference between the first and last sessions of standard drone transfer and land training, a single outlier was identified and removed. Additionally, one outlier was identified in the differences between the first session of the heavy drone transfer and land completion times and the last session of the standard drone transfer and land completion times. After removing outliers, all six Shapiro-Wilk tests had *p*-values above 0.1. All statistical analyses were conducted using *R* studio 2021.

3. Results and discussion

The result of the linear mixed models (Tables 3 and 4) showed that behavioral and biometric measures significantly affected task completion times, with roughly half (50%) of the vertical control performance profile determined by individual differences. The marginal R^2 was 0.26 and the conditional R^2 was 0.633. For the transfer and land task, Individual differences explained 12% of the performance profile variability. The marginal R^2 was 0.508 and the conditional R^2 was 0.566.

Figures 5 and 6 illustrate the main effects and the interaction effects of affective behaviors on the vertical task completion times. Figures 7 and 8 illustrate the main effect and the interaction effects of affective behaviors on the transfer and land task completion time. The existence of these relationships and that there exist differential impacts, with respect to task type and difficulty level, emphasizes the importance of factoring in the human-system interactions in drone pilot state and performance assessment (see Figure 2). In the subsections below, the results related to each research question are discussed and practical applications of those questions are presented.

Table 3. Vertical control completion time ($n = 240$ observations).

Predictors	Coefficients	CI	p
Intercept	61.38	56.4 to 66.4	**<0.001**
(1) Cognitive ability (OSPAN)	1.20	−2.93 to 5.33	0.569
(2) Cognitive load (alpha PC1)	6.70	4.11 to 9.28	**<0.001**
(3) Focused attention (theta PC1)	1.65	0.53 to 2.76	**0.004**
(4) Effort (pupil dilation)	3.38	−0.74 to 7.51	0.107
(5) Vertical guidance input	3.97	2.06 to 5.89	**<0.001**
(6) Vertical stabilization input	−3.00	−3.86 to −2.15	**<0.001**
(1) × (4)	−5.23	−8.24 to −2.21	**0.001**
Random effects (15 participants)			
σ^2	52.90		
$\tau_{00\ participant}$	53.79		
Intraclass correlation coefficient	0.50		
Marginal R^2/conditional R^2	0.260/0.633		

Bold values indicate statistical significance at the 5% level or better.

Table 4. Transfer and land completion time ($n = 240$ observations).

Predictors	Coefficients	CI	p
Intercept	75.80	67.96 to 83.64	**<0.001**
(1) Difficulty (drone weight)	−29.85	−38.81 to −20.90	**<0.001**
(2) Cognitive load (alpha PC1)	8.48	1.91 to 15.05	**0.012**
(3) Alertness (beta PC1)	2.25	−0.05 to 4.55	0.055
(4) Effort (pupil dilation)	10.86	4.35 to 17.38	**0.001**
(5) Vertical guidance input	−0.86	−1.44 to −0.27	**0.005**
(6) Sideward guidance input	−3.61	−4.53 to −2.69	**<0.001**
(7) Turn stabilization input	5.43	1.89 to 8.97	**0.003**
(8) Forward stabilization input	−9.34	−17.96 to −0.73	**0.034**
(1) × (2)	−10.70	−20.57 to −0.84	**0.034**
(1) × (3)	−3.38	−6.53 to −0.23	**0.036**
(1) × (4)	−15.07	−23.57 to −6.57	**0.001**
(1) × (6)	3.07	1.96 to 4.19	**<0.001**
(1) × (8)	11.96	1.73 to 22.19	**0.022**
Random effects (15 participants)			
σ^2	261.48		
$\tau_{00\ participant}$	35.11		
Intraclass correlation coefficient	0.12		
Marginal R^2/conditional R^2	0.508/0.566		

Bold values indicate statistical significance at the 5% level or better.

3.1. Q1: Are biometric and behavioral measures associated with drone piloting skill acquisition task-specific?

In a fashion similar to that in prior literature on drone pilot learning, reduced cognitive demands and enhanced attentional control were associated with greater performance (Figures 5–8). However, some biometric and behavioral markers associated with skill acquisition were task-specific. Specifically, effort, focused attention, and vigilance had task-dependent differential effects on performance. This may help to explain why there is a discrepancy between biometric and behavioral measures in the extant literature on drone piloting skill acquisition. It can be inferred that drone piloting involves underlying cognitive processes and requirements that differ by task. Thus, choosing the most effective measures is crucial for identifying pilots' states to design drone tasks and develop related training programs. Subsequent sections describe each of the biometric and behavioral measures tested and how they relate to two specific drone activities.

3.1.1. Cognitive load

Cognitive load, based on the first principal component of alpha power (i.e., alpha PC1), had a significant effect on vertical control task completion times (t-test, $p < 0.001$) as well as transfer and land completion times (t-test, $p = 0.012$), with higher cognitive loads associated with longer completion times. As such, drone pilot skill acquisition is closely

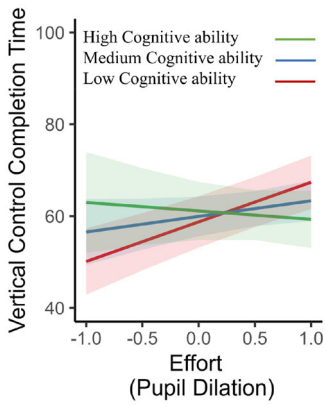

Figure 6. Vertical control completion time interaction.

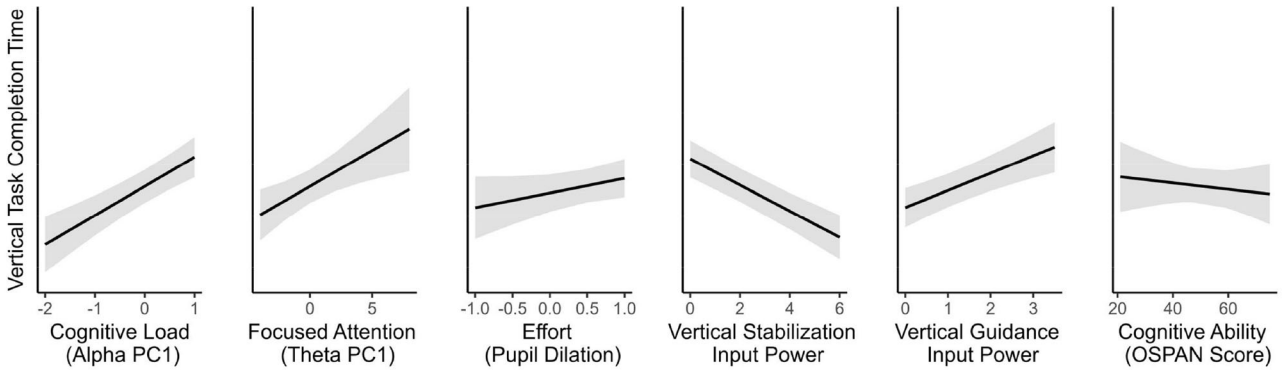

Figure 5. Vertical control completion time main effects.

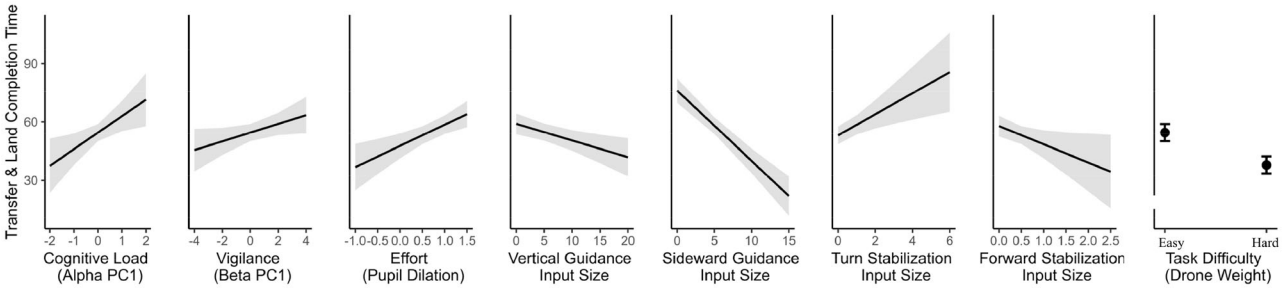

Figure 7. Transfer and land task completion time main effects.

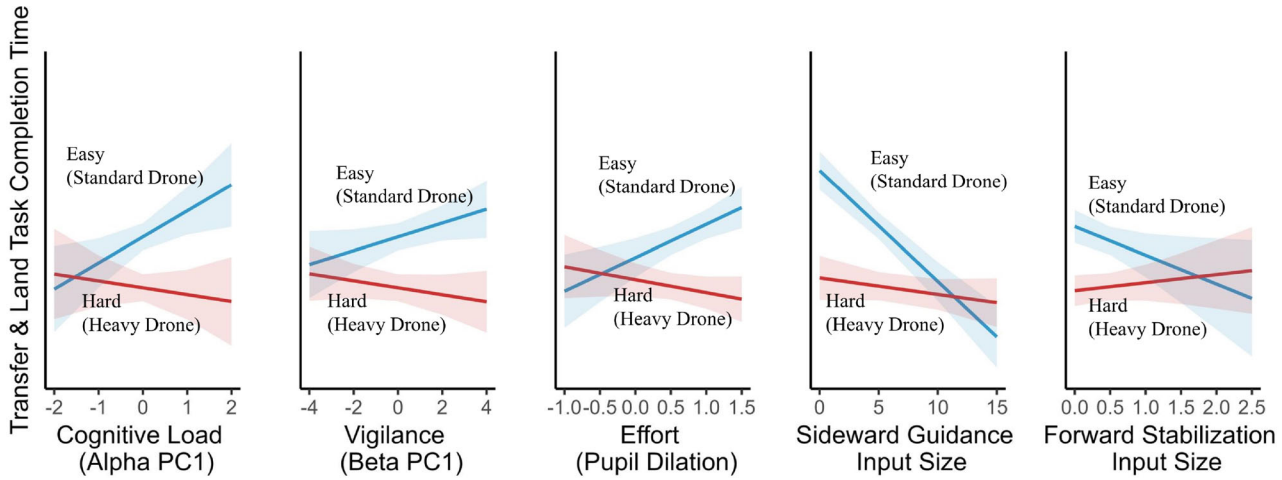

Figure 8. Transfer and land task completion time interactions.

linked with cognitive load levels. For the transfer and land task, the effect of cognitive load (alpha PC1) on completion times interacted with the task difficulty (t-test, $p = 0.034$). At the low difficulty level, greater cognitive load resulted in lesser performance, while for the high difficulty level, the cognitive load effect on completion times was not significant. This may be an indication that for this task, trainees are moving from the cognitive stage of learning to a more autonomous stage. Consequently, practice at that point no longer reduced cognitive load. It was also noted that the performance was not affected by the increased payload for the transfer and land task. An alternate explanation is that trainees use the extra cognitive resources available due to practice to increase their efficiency for the complex level of transfer and land tasks, so that the cognitive load remains the same, as the pilot becomes efficient. This is also observed in the gradual improvement in performance. When trainees are not in the cognitive stage of learning (Figure 1), but are overtraining to ensure mastery, the cognitive load measured by alpha power may not reflect the effectiveness of learning. The effectiveness of training should rather be gauged through a combination of other subjective and objective measures.

3.1.2. Focused attention

There was a significant relationship between focused attention (theta PC1) and vertical control completion times (t-test, $p = 0.004$). During skill development, participants used less focused attention to complete the vertical control tasks. However, the effect of theta PC1 (focused attention) on the transfer land task completion times was not significant. This is likely, at least in part due to fundamental differences in the control requirements for the two tasks. That is, the vertical control task was accomplished using the left stick in the vertical direction only. Both sticks in both directions (vertical, and sideways) are required to complete the transfer and landing task. The additional control requirements for transfer and land may prevent skill from reaching a level where less attentional resources would be needed to accomplish the task. Incorporating this information into training programs allows for optimizing the learning experience and tailoring it to trainees' learning stages. For example, tasks requiring fewer control requirements can be introduced earlier in the training process, while tasks requiring more complex control requirements can be introduced later in the training process and the level of each trainee's expertise can be determined based on the level of attention needed to complete the tasks.

3.1.3. Vigilance

The interaction effect for vigilance (*beta PC1*) with task complexity (difficulty) was significant for the transfer and land task completion times (t-test, $p = 0.036$). The interaction plot (Figure 8) suggests opposing effects from the two task complexities. For the standard drone greater vigilance (i.e., higher Beta PC1) was associated with longer task completion times. However, for the heavy drone, greater vigilance did not improve performance. Thus, the greater difficulty of controlling the heavy drone may be challenging

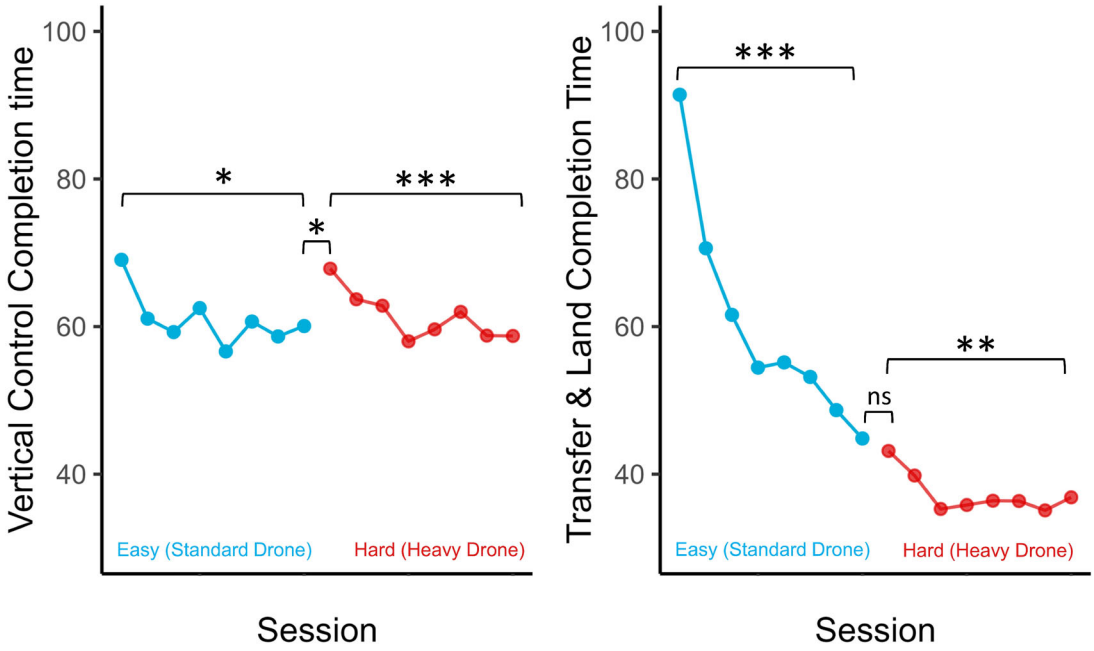

Figure 9. Task completion times for vertical control and transfer and land tasks.

enough that increased vigilance does not have a positive effect and may actually hinder performance. For the standard drone, pilots may become more vigilant when stressed by complex drone missions, but that vigilance may not lead to superior performance. Thus, the level of vigilance provides potentially valuable information for understanding efficiency training programs and promotes competency and accuracy when performing the assigned tasks. For the vertical control task, the effect of vigilance on completion times was not significant. This task dependent result might be due to fundamental differences in the underlying cognitive processes needed for maintaining a specific altitude *vs.* moving to a specific location and landing.

3.2. Q2: Does cognitive ability moderate the effect of biometric measures of performance?

There are several significant behavioral effects that influence performance for both tasks. For the vertical control task, *cognitive load*, *focused attention*, and *effort* each influenced performance. For the transfer and land task, *cognitive load*, *alertness*, and *effort* each influenced performance. The cognitive load effect was similar in magnitude for both tasks, however, the effect of effort was somewhat stronger for the transfer and land task, suggesting that it was somewhat more challenging generally.

With respect to interaction/moderation effects, for the vertical control task, the effect of effort (as measured by pupillary diameter) was moderated by *cognitive ability* (Figure 6 and Table 3—Interaction effect, t-test, $p < 0.001$). Participants with greater cognitive ability performed better while exerting greater effort. Notably, for participants with lesser cognitive ability, greater effort was associated with lesser performance. For the transfer and land task, cognitive ability was not significant as a main effect nor as a moderating (interaction) effect. These findings highlight the value of quantifying individual differences, such as cognitive ability.

3.3. Q3: Does the amount of control input behavior impact performance?

A two-level model of controller input behavior was used for assessing expertise. The vertical stabilization input was larger during vertical control skill development (t-test, $p < 0.001$), and the vertical guidance input was correspondingly smaller (t-test, $p < 0.001$). Intuitively, greater stabilization power might improve performance. However, results showed that greater vertical guidance input indicated flight excursions that resulted in lesser performance.

Larger vertical guidance inputs (t-test, $p = 0.005$) and smaller turn stabilization inputs (t-test, $p = 0.003$) for the transfer and land task were associated with better performance. Also, larger sideward guidance inputs led to better performance (t-test, $p < 0.001$). This effect was greater at low difficulty levels (t-test, $p < 0.001$). Similarly, larger forward stabilization inputs led to better performance (t-test, $p = 0.034$). But the effect of forward stabilization input size on transfer and land task completion time was dependent on the difficulty level of the task (t-test, $p = 0.022$). Larger forward stabilization inputs were associated with shorter completion times at the low difficulty level, but longer completion times at high difficulty.

As a result, there is support for employing controller input behavior in user-centered design for performance. Data from controller input provides continuous, inexpensive, non-intrusive, and robust information about pilot performance. The design of new features, advanced systems, and safety overrides can benefit from understanding and considering pilots' control responses. For example, during a handoff of an assistive system, instability-induced control

behaviors may suggest alternative designs. Similarly, a change in drone dynamic parameters, such as controller sensitivity that may lead to dynamic instability or suboptimal mental states, might be avoided. Such information can be used to provide pilot feedback such that the trainee can focus their practice on maneuvers that involve instability-induced control behaviors.

3.4. Q4: How does increased task difficulty (weight) affect task performance?

Continued practice tended to improve performance on both tasks and at both difficulty levels. However, the effects of greater task difficulty varied between the two tasks. Figure 9 illustrates the completion times for vertical and transfer and land tasks at each difficulty level. Completion times for both tasks improved from the first to the last sessions of practice using the standard drone (paired *t*-test, vertical control $p = 0.026$, transfer and land $p < 0.001$). The task completion times for the more difficult (heavy drone) treatment also improved between the first and last practice sessions (paired *t*-test, vertical control $p < 0.001$, transfer and land $p < 0.01$). Changing to higher difficulty, however, affected the learning process differentially for the two tasks. For vertical control, performance dropped (paired *t*-test, $p = 0.01$) suggesting that the task had to be relearned with additional practice. The difference in transfer and land task performance between the last session of practice with the standard drone and the first session of practice with the heavy drone was not significant (paired *t*-test, $p = 0.249$).

The amount of relearning required due to an increase in task difficulty varied by task type. Heuristic evaluations of drone designs or usability tests on a single mission may not provide sufficient information for optimal human-in-the-loop system design. That is, design usability should be evaluated across a variety of drone missions. A training regimen might follow significant design change, focusing on relearning or retraining specific tasks. However, this may require the identification of specific drone flight requirements and technologies. Drone missions might be decomposed into specific sub-tasks, with each evaluated for usability when designing a drone/pilot system interface.

3.5. Limitations

Experiments were conducted in a controlled environment to minimize distractors and to optimize statistical power in the context of measuring, modeling, and understanding human performance. While the value of simulations is widely recognized, follow-on field study to verify current findings in practical or operational settings will be valuable. This study centered on two critical early phase skills, using novice pilots, for drone pilot skill acquisition, namely vertical control, and transfer and land. Professional drone pilots come from diverse backgrounds and have a potentially broad range of skill, experience, education, and training, which can affect their behavior and performance in varied situations. Thus, care must be taken in generalizing toward the broader range of *a priori* skills that exist in an industry that is rapidly evolving, adding new technologies, operational practices, and regulatory requirements. To explore the application of proposed measures and recommendations to refine industry-specific drone operations, further research will be valuable in addressing such generalization.

4. Conclusions

The development of drone assistive technologies and the ever-increasing use of drones for civilian purposes raises interesting questions regarding pilots' interactions with drones and pilot training. The modeling of pilots' control actions and mental states, representing such interactions, is essential for the successful adaptation of drones in industry 4.0. This study identified biometrics and behavioral indicators for drone-pilot skill learning across tasks and difficulty levels. A potential next step could be to develop AI-based training paradigms that incorporate these behavioral and biometric factors as inputs to create engaging and tailored learning experiences. In addition, these learning paradigms can be incorporated into simulation-based learning platforms to enable the creation of complex scenarios that represent drone missions more accurately. Pilots would then be able to gain a deeper understanding of how to handle the range of complexities in real-world drone operations without the costs, risks, and dangers associated with real-life drone operations.

Organizations should invest in job design and usability studies to ensure that drone operations are safe and effective. This study provided recommendations and methods for testing drone operation usability and learnability, finding that behavioral and biometric metrics for assessing usability and learnability testing should be selected based on factors including skill level, individual differences in cognitive ability, task type, and task difficulty. Having multiple affective measures, in turn, allows for addressing a broader range of drone-operation issues and identifying nuanced differences in pilots' performance in various flight situations and environments. Future research will aid in the development of more advanced and intuitive user interfaces for drone operations.

Disclosure statement

No potential conflict of interest was reported by the author(s).

ORCID

Fatemeh Dalilian http://orcid.org/0000-0002-4173-3345
David Nembhard http://orcid.org/0000-0003-2262-6840

References

Abdelrahman, Y., Velloso, E., Dingler, T., Schmidt, A., & Vetere, F. (2017). Cognitive heat: Exploring the usage of thermal imaging to unobtrusively estimate cognitive load. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, *1*(3), 1–20. https://doi.org/10.1145/3130898

Ahern, S., & Beatty, J. (1979). Pupillary responses during information processing vary with Scholastic Aptitude Test scores. *Science*, *205*(4412), 1289–1292. https://doi.org/10.1126/science.472746

Albeaino, G., Eiris, R., Gheisari, M., & Issa, R. R. (2022). DroneSim: A VR-based flight training simulator for drone-mediated building inspections. *Construction Innovation*, *22*(4), 831–848. https://doi.org/10.1108/CI-03-2021-0049

Andreassi, J. L. (2010). *Psychophysiology: Human behavior and physiological response*. Psychology Press.

Astolfi, L., Cincotti, F., Babiloni, C., Carducci, F., Basilisco, A., Rossini, P. M., Salinari, S., Mattia, D., Cerutti, S., Dayan, D. B., Ding, L., Ni, Y., He, B., & Babiloni, F. (2005). Estimation of the cortical connectivity by high-resolution EEG and structural equation modeling: Simulations and application to finger tapping data. *IEEE Transactions on Bio-Medical Engineering*, *52*(5), 757–768. https://doi.org/10.1109/TBME.2005.845371

Backs, R. W., & Walrath, L. C. (1992). Eye movement and pupillary response indices of mental workload during visual search of symbolic displays. *Applied Ergonomics*, *23*(4), 243–254. https://doi.org/10.1016/0003-6870(92)90152-l

Baddeley, A. D., & Hitch, G. (1974). *Working memory (Psychology of learning and motivation)* (Vol. 8, pp. 47–89). Elsevier.

Basile, L. F., Anghinah, R., Ribeiro, P., Ramos, R. T., Piedade, R., Ballester, G., & Brunetti, E. P. (2007). Interindividual variability in EEG correlates of attention and limits of functional mapping. *International Journal of Psychophysiology*, *65*(3), 238–251. https://doi.org/10.1016/j.ijpsycho.2007.05.001

Berka, C., Levendowski, D. J., Westbrook, P., Davis, G., Lumicao, M. N., Olmstead, R. E., Popovic, M., Zivkovic, V. T., & Ramsey, C. K. (2005). EEG quantification of alertness: Methods for early identification of individuals most susceptible to sleep deprivation. In *Biomonitoring for Physiological and Cognitive Performance during Military Operations*. https://doi.org/10.1117/12.597503

Besserve, M., Philippe, M., Florence, G., Laurent, F., Garnero, L., & Martinerie, J. (2008). Prediction of performance level during a cognitive task from ongoing EEG oscillatory activities. *Clinical Neurophysiology*, *119*(4), 897–908. https://doi.org/10.1016/j.clinph.2007.12.003

Bo, J., Borza, V., & Seidler, R. D. (2009). Age-related declines in visuospatial working memory correlate with deficits in explicit motor sequence learning. *Journal of Neurophysiology*, *102*(5), 2744–2754. https://doi.org/10.1152/jn.00393.2009

Brown, S., & Bennett, E. (2002). The role of practice and automaticity in temporal and nontemporal dual-task performance. *Psychological Research*, *66*(1), 80–89. https://doi.org/10.1007/s004260100076

Buszard, T., Farrow, D., Verswijveren, S. J., Reid, M., Williams, J., Polman, R., Ling, F. C. M., & Masters, R. S. (2017). Working memory capacity limits motor learning when implementing multiple instructions. *Frontiers in Psychology*, *8*, 1350. https://doi.org/10.3389/fpsyg.2017.01350

Cardona-Reyes, H., Trujillo-Espinoza, C., Arevalo-Mercado, C., & Muñoz-Arteaga, J. (2021). Training of drone pilots through virtual reality environments under the gamification approach in a university context.

Chen, D., Wang, M., He, C., Luo, Q., Iravantchi, Y., Sample, A., Shin, K. G., & Wang, X. (2021). MagX: Wearable, untethered hands tracking with passive magnets. In *Proceedings of the 27th Annual International Conference on Mobile Computing and Networking*. https://doi.org/10.1145/3447993.3511175

Cole, D. T., Thompson, P., Göktoğan, A. H., & Sukkarieh, S. (2010). System development and demonstration of a cooperative UAV team for mapping and tracking. *The International Journal of Robotics Research*, *29*(11), 1371–1399. https://doi.org/10.1177/0278364910364685

Comolli, F., Gobbi, M., & Mastinu, G. (2020). Study on the driver/steering wheel interaction in emergency situations. *Applied Sciences*, *10*(20), 7055. https://doi.org/10.3390/app10207055

Congress, S. S., Puppala, A. J., & Lundberg, C. L. (2018). Total system error analysis of UAV-CRP technology for monitoring transportation infrastructure assets. *Engineering Geology*, *247*, 104–116. https://doi.org/10.1016/j.enggeo.2018.11.002

Conway, A. R., Cowan, N., Bunting, M. F., Therriault, D. J., & Minkoff, S. R. (2002). A latent variable analysis of working memory capacity, short-term memory capacity, processing speed, and general fluid intelligence. *Intelligence*, *30*(2), 163–183. https://doi.org/10.1016/S0160-2896(01)00096-4

Corichi, E. Z., Carranza, J. M., Garca, C. A. R., & Pineda, L. V. (2017). Real-time prediction of altered states in Drone pilots using physiological signals. In *2017 Workshop on Research, Education and Development of Unmanned Aerial Systems (RED-UAS)*. https://doi.org/10.1109/RED-UAS.2017.8101674

Crandall, J. W., Goodrich, M. A., Olsen, D. R., & Nielsen, C. W. (2005). Validating human–robot interaction schemes in multitasking environments. *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, *35*(4), 438–449. https://doi.org/10.1109/TSMCA.2005.850587

Dalenogare, L. S., Benitez, G. B., Ayala, N. F., & Frank, A. G. (2018). The expected contribution of Industry 4.0 technologies for industrial performance. *International Journal of Production Economics*, *204*, 383–394. https://doi.org/10.1016/j.ijpe.2018.08.019

De Cecilio de Carlos, R. (2019). *Pupil size manipulation in covert attention to red and blue contrasts: Study and application in a human computer interface*.

De Gennaro, L., Ferrara, M., & Bertini, M. (2001). The boundary between wakefulness and sleep: Quantitative electroencephalographic changes during the sleep onset period. *Neuroscience*, *107*(1), 1–11. https://doi.org/10.1016/s0306-4522(01)00309-8

Donges, E. (1978). A two-level model of driver steering behavior. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *20*(6), 691–707. https://doi.org/10.1177/001872087802000607

Du, X., Zhang, Y., Tian, Y., Huang, W., Wu, B., & Zhang, J. (2015). The influence of spatial ability and experience on performance during spaceship rendezvous and docking. *Frontiers in Psychology*, *6*, 955. https://doi.org/10.3389/fpsyg.2015.00955

Economic Report (2017). Association for Unmanned Vehicle Systems International. Retrieved August 26, from https://www.auvsi.org/our-impact/economic-report

Eiris, R., Albeaino, G., Gheisari, M., Benda, W., & Faris, R. (2021). InDrone: A 2D-based drone flight behavior visualization platform for indoor building inspection. *Smart and Sustainable Built Environment*, *10*(3), 438–456. https://doi.org/10.1108/SASBE-03-2021-0036

Engle, R. W. (2002). Working memory capacity as executive attention. *Current Directions in Psychological Science*, *11*(1), 19–23. https://doi.org/10.1111/1467-8721.00160

Farina, D., Stegeman, D., & Merletti, R. (2016). Biophysics of the generation of EMG signals. *Surface Electromyography: Physiology, Engineering, and Applications*, 1–24.

Fitts, P. M., & Posner, M. I. (1967). *Human performance*. Brooks/Cole.

Fletcher, K., Neal, A., & Yeo, G. (2017). The effect of motor task precision on pupil diameter. *Applied Ergonomics*, *65*, 309–315. https://doi.org/10.1016/j.apergo.2017.07.010

Fong, W., Ong, S., & Nee, A. (2008). Methods for in-field user calibration of an inertial measurement unit without external equipment. *Measurement Science and Technology*, *19*(8), 085202. https://doi.org/10.1088/0957-0233/19/8/085202

Forte, G., Favieri, F., & Casagrande, M. (2019). Heart rate variability and cognitive function: A systematic review. *Frontiers in Neuroscience*, *13*, 710. https://doi.org/10.3389/fnins.2019.00710

Furley, P. A., & Memmert, D. (2010). The role of working memory in sport. *International Review of Sport and Exercise Psychology*, *3*(2), 171–194. https://doi.org/10.1080/1750984X.2010.526238

Gevins, A., & Smith, M. E. (2000). Neurophysiological measures of working memory and individual differences in cognitive ability and cognitive style. *Cerebral Cortex*, *10*(9), 829–839. https://doi.org/10.1093/cercor/10.9.829

Girden, E. R. (1992). *ANOVA: Repeated measures*. Sage.

Gonzalez, C. (2005). Task workload and cognitive abilities in dynamic decision making. *Human Factors*, *47*(1), 92–101. https://doi.org/10.1518/0018720053653767

Goodie, J. L., Larkin, K. T., & Schauss, S. (2000). Validation of Polar heart rate monitor for assessing heart rate during physical and mental stress. *Journal of Psychophysiology*, *14*(3), 159–164. https://doi.org/10.1027//0269-8803.14.3.159

Gu, H., Yao, Q., Chen, H., Ding, Z., Zhao, X., Liu, H., Feng, Y., Li, C., & Li, X. (2022). The effect of mental schema evolution on mental workload measurement: An EEG study with simulated quadrotor UAV operation. *Journal of Neural Engineering*, *19*(2), 026058. https://doi.org/10.1088/1741-2552/ac6828

Hajcak, G., MacNamara, A., & Olvet, D. M. (2010). Event-related potentials, emotion, and emotion regulation: An integrative review. *Developmental Neuropsychology*, *35*(2), 129–155. https://doi.org/10.1080/87565640903526504

Hess, R., & Modjtahedzadeh, A. (1990). A control theoretic model of driver steering behavior. *IEEE Control Systems Magazine*, *10*(5), 3–8. https://doi.org/10.1109/37.60415

Hultman, E., & Sjöholm, H. (1983). Electromyogram, force and relaxation time during and after continuous electrical stimulation of human skeletal muscle *in situ*. *The Journal of Physiology*, *339*(1), 33–40. https://doi.org/10.1113/jphysiol.1983.sp014700

Intelligence, I. (2021). *Drone technology uses and applications for commercial, industrial and military drones in 2021 and the future*. Business Insider. Retrieved August 26, from https://www.businessinsider.com/drone-technology-uses-applications

Ishii, R., Canuet, L., Ishihara, T., Aoki, Y., Ikeda, S., Hata, M., Katsimichas, T., Gunji, A., Takahashi, H., Nakahachi, T., Iwase, M., & Takeda, M. (2014). Frontal midline theta rhythm and gamma power changes during focused attention on mental calculation: An MEG beamformer analysis. *Frontiers in Human Neuroscience*, *8*, 406. https://doi.org/10.3389/fnhum.2014.00406

Jao, P.-K., Chavarriaga, R., Dell'Agnola, F., Arza, A., Atienza, D., & Millan, J. d R. (2021). EEG correlates of difficulty levels in dynamical transitions of simulated flying and mapping tasks. *IEEE Transactions on Human–Machine Systems*, *51*(2), 99–108. https://doi.org/10.1109/THMS.2020.3038339

Jaquess, K. J., Lo, L.-C., Oh, H., Lu, C., Ginsberg, A., Tan, Y. Y., Lohse, K. R., Miller, M. W., Hatfield, B. D., & Gentili, R. J. (2018). Changes in mental workload and motor performance throughout multiple practice sessions under various levels of task difficulty. *Neuroscience*, *393*, 305–318. https://doi.org/10.1016/j.neuroscience.2018.09.019

Javaid, M., Khan, I. H., Singh, R. P., Rab, S., & Suman, R. (2022). Exploring contributions of drones towards Industry 4.0. *Industrial Robot: The International Journal of Robotics Research and Application*, *49*(3), 476–490. https://doi.org/10.1108/IR-09-2021-0203

Kaczorowska, M., Plechawska-Wojcik, M., Tokovarov, M., Dmytruk, R. (2017). *Comparison of the ICA and PCA methods in correction of EEG signal artefacts* [Paper presentation]. 2017 10th International Symposium on Advanced Topics in Electrical Engineering (ATEE). https://doi.org/10.1109/ATEE.2017.7905095

Kadir, B. A., Broberg, O., & da Conceicao, C. S. (2019). Current research and future perspectives on human factors and ergonomics in Industry 4.0. *Computers & Industrial Engineering*, *137*, 106004. https://doi.org/10.1016/j.cie.2019.106004

Kamiński, J., Brzezicka, A., Gola, M., & Wróbel, A. (2012). Beta band oscillations engagement in human alertness process. *International Journal of Psychophysiology*, *85*(1), 125–128. https://doi.org/10.1016/j.ijpsycho.2011.11.006

Kang, H., Li, H., Zhang, J., Lu, X., & Benes, B. (2018). Flycam: Multitouch gesture controlled drone gimbal photography. *IEEE Robotics and Automation Letters*, *3*(4), 3717–3724. https://doi.org/10.1109/LRA.2018.2856271

Kerick, S. E., Douglass, L. W., & Hatfield, B. D. (2004). Cerebral cortical adaptations associated with visuomotor practice. *Medicine and Science in Sports and Exercise*, *36*(1), 118–129. https://doi.org/10.1249/01.MSS.0000106176.31784.D4

Kong, F. (2019). Development of metric method and framework model of integrated complexity evaluations of production process for ergonomics workstations. *International Journal of Production Research*, *57*(8), 2429–2445. https://doi.org/10.1080/00207543.2018.1519266

Kopp, T., Baumgartner, M., & Kinkel, S. (2021). Success factors for introducing industrial human–robot interaction in practice: An empirically driven framework. *The International Journal of Advanced Manufacturing Technology*, *112*(3–4), 685–704. https://doi.org/10.1007/s00170-020-06398-0

Kramer, M. (2004). *Automatic model selection in the mixed models framework* [Paper presentation]. Proceedings, 16th Annual Kansas State University Conference on Applied Statistics in Agriculture. https://doi.org/10.4148/2475-7772.1155

Krejtz, K., Duchowski, A. T., Niedzielska, A., Biele, C., & Krejtz, I. (2018). Eye tracking cognitive load using pupil diameter and microsaccades with fixed gaze. *PLOS One*, *13*(9), e0203629. https://doi.org/10.1371/journal.pone.0203629

Kruijff, G.-J M., Pirri, F., Gianni, M., Papadakis, P., Pizzoli, M., Sinha, A., Tretyakov, V., Linder, T., Pianese, E., & Corrao, S. (2012). Rescue robots at earthquake-hit Mirandola, Italy: A field report. In *2012 IEEE International Symposium on Safety, Security, and Rescue Robotics (SSRR)*. https://doi.org/10.1109/SSRR.2012.6523866

Land, M. F. (1998). The visual control of steering. *Vision and Action*, *28*, 168–180.

Lehtonen, E., Lappi, O., Koirikivi, I., & Summala, H. (2014). Effect of driving experience on anticipatory look-ahead fixations in real curve driving. *Accident; Analysis and Prevention*, *70*, 195–208. https://doi.org/10.1016/j.aap.2014.04.002

Li, L. (2022). Reskilling and upskilling the future-ready workforce for Industry 4.0 and beyond. *Information Systems Frontiers*, 1–16. https://doi.org/10.1007/s10796-022-10308-y

Li, L., Chen, R., & Chen, J. (2016). Playing action video games improves visuomotor control. *Psychological Science*, *27*(8), 1092–1108. https://doi.org/10.1177/0956797616650300

Li, L., Sweet, B. T., & Stone, L. S. (2005). Effect of contrast on the active control of a moving line. *Journal of Neurophysiology*, *93*(5), 2873–2886. https://doi.org/10.1152/jn.00200.2004

Liu, Y., & Wickens, C. D. (1994). Mental workload and cognitive task automaticity: An evaluation of subjective and time estimation metrics. *Ergonomics*, *37*(11), 1843–1854. https://doi.org/10.1080/00140139408964953

Lohani, M., Payne, B. R., & Strayer, D. L. (2019). A review of psychophysiological measures to assess cognitive states in real-world driving. *Frontiers in Human Neuroscience*, *13*, 57. https://doi.org/10.3389/fnhum.2019.00057

Luinge, H. J., & Veltink, P. H. (2005). Measuring orientation of human body segments using miniature gyroscopes and accelerometers. *Medical & Biological Engineering & Computing*, *43*(2), 273–282. https://doi.org/10.1007/BF02345966

Manka, M. (2022). Developing an efficient real-time terrestrial infrastructure inspection system using autonomous drones and deep learning.

Markkula, G., Benderius, O., & Wahde, M. (2014). Comparing and validating models of driver steering behaviour in collision avoidance and vehicle stabilisation. *Vehicle System Dynamics*, *52*(12), 1658–1680. https://doi.org/10.1080/00423114.2014.954589

MatLab, P. (2020). *9.9. 0.1495850 (R2020b)*. The MathWorks Inc.

Maza, I., Caballero, F., Molina, R., Peña, N., & Ollero, A. (2010). Multimodal interface technologies for UAV ground control stations: A comparative analysis. In *Selected Papers from the 2nd International Symposium on UAVs, Reno, Nevada, USA June 8–10, 2009* (pp. 371–391). Springer.

Merceron, A., Blikstein, P., & Siemens, G. (2016). Learning analytics: From big data to meaningful data. *Journal of Learning Analytics*, *2*(3), 4–8. https://doi.org/10.18608/jla.2015.23.2

Mole, C. D., Kountouriotis, G., Billington, J., & Wilkie, R. M. (2016). Optic flow speed modulates guidance level control: New insights into two-level steering. *Journal of Experimental Psychology: Human Perception and Performance*, *42*(11), 1818. https://doi.org/10.1037/xhp0000256

Mu, S., Cui, M., & Huang, X. (2020). Multimodal data fusion in learning analytics: A systematic review. *Sensors*, *20*(23), 6856. https://doi.org/10.3390/s20236856

Nakano, H., Osumi, M., Ueta, K., Kodama, T., & Morioka, S. (2013). Changes in electroencephalographic activity during observation, preparation, and execution of a motor learning task. *The International Journal of Neuroscience*, *123*(12), 866–875. https://doi.org/10.3109/00207454.2013.813509

Olivares-Figueroa, J. D., Cruz-Vega, I., Ramírez-Cortés, J., Gómez-Gil, P., & Martínez-Carranza, J. (2021). A compact approach for emotional assessment of drone pilots using BCI.

Pacaux-Lemoine, M.-P., & Trentesaux, D. (2019). Ethical risks of human–machine symbiosis in industry 4.0: Insights from the human–machine cooperation approach. *IFAC-PapersOnLine*, *52*(19), 19–24. https://doi.org/10.1016/j.ifacol.2019.12.077

Park, J., Park, J., Shin, D., & Choi, Y. (2021). A BCI based alerting system for attention recovery of UAV operators. *Sensors*, *21*(7), 2447. https://doi.org/10.3390/s21072447

Peißl, S., Wickens, C. D., & Baruah, R. (2018). Eye-tracking measures in aviation: A selective literature review. *The International Journal of Aerospace Psychology*, *28*(3–4), 98–112. https://doi.org/10.1080/24721840.2018.1514978

Peng, X., Stone, L. S., & Li, L. (2010). Humans can control heading independent of visual path information. *Journal of Vision*, *8*(6), 1160–1160. https://doi.org/10.1167/8.6.1160

Peysakhovich, V., Vachon, F., Vallières, B. R., Dehais, F., & Tremblay, S. (2015). Pupil dilation and eye movements can reveal upcoming choice in dynamic decision-making. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*. https://doi.org/10.1177/1541931215591043

Pfeiffer, C., & Scaramuzza, D. (2021). Human-piloted drone racing: Visual processing and control. *IEEE Robotics and Automation Letters*, *6*(2), 3467–3474. https://doi.org/10.1109/LRA.2021.3064282

Pham, T., Tezza, D., & Andujar, M. (2020). Enhancing drone pilots' engagement through a brain-computer interface. In *International Conference on Human–Computer Interaction*.

Plazak, J., DiGiovanni, D. A., Collins, D. L., & Kersten-Oertel, M. (2019). Cognitive load associations when utilizing auditory display within image-guided neurosurgery. *International Journal of Computer Assisted Radiology and Surgery*, *14*(8), 1431–1438. https://doi.org/10.1007/s11548-019-01970-w

Poldrack, R. A., Sabb, F. W., Foerde, K., Tom, S. M., Asarnow, R. F., Bookheimer, S. Y., & Knowlton, B. J. (2005). The neural correlates of motor skill automaticity. *The Journal of Neuroscience*, *25*(22), 5356–5364. https://doi.org/10.1523/JNEUROSCI.3880-04.2005

Polfreman, R. (2018). Hand posture recognition: IR, sEMG and IMU. In *Proceedings of the Conference on New Interfaces for Musical Expression 2018*, Blacksburg, VA, 6.

Posner, M. I. (1975). Psychobiology of attention. In M. S. Gazzaniga & C. Blakemore (Eds.), *Handbook of psychobiology* (pp. 441–480). Academic Press.

Qiao, H., Tahara, A., Nakphu, N., & Iramina, K. (2022). Using mean pupil diameter change to analyze behavioral performance in multitasking training game. In *2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*. https://doi.org/10.1109/EMBC48229.2022.9871751

Ramautar, J. R., Romeijn, N., Gómez-Herrero, G., Piantoni, G., & Van Someren, E. J. (2013). Coupling of infraslow fluctuations in autonomic and central vigilance markers: Skin temperature, EEG beta power and ERP P300 latency. *International Journal of Psychophysiology*, *89*(2), 158–164. https://doi.org/10.1016/j.ijpsycho.2013.01.001

Razafimahazo, E., De Saqui-Sannes, P., Vingerhoeds, R. A., Baron, C., Soulax, J., & Mège, R. (2021). Mastering complexity for indoor inspection drone development. in *2021 IEEE International Symposium on Systems Engineering (ISSE)*. https://doi.org/10.1109/ISSE51541.2021.9582483

Ribeiro, R., Ramos, J., Safadinho, D., Reis, A., Rabadão, C., Barroso, J., & Pereira, A. (2021). Web AR solution for UAV pilot training and usability testing. *Sensors*, *21*(4), 1456. https://doi.org/10.3390/s21041456

Rioult-Pedotti, M.-S., Friedman, D., & Donoghue, J. P. (2000). Learning-induced LTP in neocortex. *Science*, *290*(5491), 533–536. https://doi.org/10.1126/science.290.5491.533

Rizzi, E., Jagacinski, R. J., & Bloom, B. J. (2021). Spatio-temporal flexibility of attention inferred from drivers' steering movements. *Journal of Motor Behavior*, *53*(6), 758–769. https://doi.org/10.1080/00222895.2020.1868968

Rode, C., Robson, R., Purviance, A., Geary, D. C., & Mayr, U. (2014). Is working memory training effective? A study in a school setting. *PLOS One*, *9*(8), e104796. https://doi.org/10.1371/journal.pone.0104796

Rohde, T. E., & Thompson, L. A. (2007). Predicting academic achievement with cognitive ability. *Intelligence*, *35*(1), 83–92. https://doi.org/10.1016/j.intell.2006.05.004

Roldán-Gómez, J. J., González-Gironda, E., & Barrientos, A. (2021). A survey on robotic technologies for forest firefighting: Applying drone swarms to improve firefighters' efficiency and safety. *Applied Sciences*, *11*(1), 363. https://doi.org/10.3390/app11010363

Roy, R. N., Bovo, A., Gateau, T., Dehais, F., & Chanel, C. P. C. (2016). Operator engagement during prolonged simulated UAV operation. *IFAC-PapersOnLine*, *49*(32), 171–176. https://doi.org/10.1016/j.ifacol.2016.12.209

Rupp, M. A., Oppold, P., & McConnell, D. S. (2013). Comparing the performance, workload, and usability of a gamepad and joystick in a complex task. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*. https://doi.org/10.1177/1541931213571398

Sakib, M. N., Chaspari, T., & Behzadan, A. H. (2021). Physiological data models to understand the effectiveness of drone operation training in immersive virtual reality. *Journal of Computing in Civil Engineering*, *35*(1), 04020053. https://doi.org/10.1061/(ASCE)CP.1943-5487.0000941

Salvucci, D. D., & Gray, R. (2004). A two-point visual control model of steering. *Perception*, *33*(10), 1233–1248. https://doi.org/10.1068/p5343

Sano, A., Taylor, S., McHill, A. W., Phillips, A. J., Barger, L. K., Klerman, E., & Picard, R. (2018). Identifying objective physiological markers and modifiable behaviors for self-reported stress and mental health status using wearable sensors and mobile phones: Observational study. *Journal of Medical Internet Research*, *20*(6), e210. https://doi.org/10.2196/jmir.9410

Schmidt, R., Schadow, J., Eißfeldt, H., & Pecena, Y. (2022). Insights on remote pilot competences and training needs of civil drone pilots. *Transportation Research Procedia*, *66*, 1–7. https://doi.org/10.1016/j.trpro.2022.12.001

Schmidt, E. A., Schrauf, M., Simon, M., Fritzsche, M., Buchner, A., & Kincses, W. E. (2009). Drivers' misjudgement of vigilance state during prolonged monotonous daytime driving. *Accident; Analysis and Prevention*, *41*(5), 1087–1093. https://doi.org/10.1016/j.aap.2009.06.007

Schwab, K., & Zahidi, S. (2020). *The future of jobs report 2020*. Retrieved from https://www3.weforum.org/docs/WEF_Future_of_Jobs_2020.pdf

Senoussi, M., Verdiere, K. J., Bovo, A., Chanel, C. P. C., Dehais, F., & Roy, R. N. (2017). Pre-stimulus antero-posterior EEG connectivity predicts performance in a UAV monitoring task. In *2017 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*.

Sheikholeslami, C., Yuan, H., He, E. J., Bai, X., Yang, L., & He, B. (2007). A high resolution EEG study of dynamic brain activity during video game play. In *2007 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. https://doi.org/10.1109/IEMBS.2007.4352833

Shipstead, Z., Redick, T. S., & Engle, R. W. (2012). Is working memory training effective? *Psychological Bulletin*, *138*(4), 628–654. https://doi.org/10.1037/a0027473

Sibley, C., Coyne, J., & Baldwin, C. (2011). Pupil dilation as an index of learning. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*. https://doi.org/10.1177/1071181311551049

Sibley, C., Coyne, J., Cole, A., Gibson, G., Baldwin, C., Roberts, D., & Barrow, J. (2010). Adaptive training in an unmanned aerial vehicle: Examination of several candidate realtime metrics. In *Applied human factors and ergonomics*. Taylor & Francis.

Smith, M. E., McEvoy, L. K., & Gevins, A. (1999). Neurophysiological indices of strategy development and skill acquisition. *Brain Research. Cognitive Brain Research*, *7*(3), 389–404. https://doi.org/10.1016/s0926-6410(98)00043-3

Tanaka, H., Hayashi, M., & Hori, T. (1996). Statistical features of hypnagogic EEG measured by a new scoring system. *Sleep*, *19*(9), 731–738. https://doi.org/10.1093/sleep/19.9.731

Tanaka, H., Hayashi, M., & Hori, T. (1997). Topographical characteristics and principal component structure of the hypnagogic EEG. *Sleep*, *20*(7), 523–534. https://doi.org/10.1093/sleep/20.7.523

Tezza, D., Laesker, D., & Andujar, M. (2021). The learning experience of becoming a FPV drone pilot. In *Companion of the 2021 ACM/IEEE International Conference on Human–Robot Interaction*. https://doi.org/10.1145/3434074.3447167

Tichon, J. G., Mavin, T., Wallis, G., Visser, T. A., & Riek, S. (2014). Using pupillometry and electromyography to track positive and negative affect during flight simulation. *Aviation Psychology and Applied Human Factors*, *4*(1), 23–32. https://doi.org/10.1027/2192-0923/a000052

TobiiConnect (2022). *How to calibrate and validate in Tobii Pro Lab*. Retrieved from https://connect.tobii.com/s/article/how-to-calibrate-and-validate-in-tobii-pro-lab?language=en_US

Travis, F., & Shear, J. (2010). Focused attention, open monitoring and automatic self-transcending: Categories to organize meditations from Vedic, Buddhist and Chinese traditions. *Consciousness and Cognition*, *19*(4), 1110–1118. https://doi.org/10.1016/j.concog.2010.01.007

Unsworth, N., Heitz, R. P., Schrock, J. C., & Engle, R. W. (2005). An automated version of the operation span task. *Behavior Research Methods*, *37*(3), 498–505. https://doi.org/10.3758/bf03192720

Van de Pol, M., & Wright, J. (2009). A simple method for distinguishing within-versus between-subject effects using mixed models. *Animal Behaviour*, *77*(3), 753–758. https://doi.org/10.1016/j.anbehav.2008.11.006

van der Wel, P., & van Steenbergen, H. (2018). Pupil dilation as an index of effort in cognitive control tasks: A review. *Psychonomic Bulletin & Review*, *25*(6), 2005–2015. https://doi.org/10.3758/s13423-018-1432-y

Van Dongen, H. P., Olofsen, E., Dinges, D. F., & Maislin, G. (2004). Mixed-model regression analysis and dealing with interindividual differences. *Methods in Enzymology*, *384*, 139–171. https://doi.org/10.1016/S0076-6879(04)84010-2

van Winsum, W., De Waard, D., & Brookhuis, K. A. (1999). Lane change manoeuvres and safety margins. *Transportation Research Part F: Traffic Psychology and Behaviour*, *2*(3), 139–149. https://doi.org/10.1016/S1369-8478(99)00011-X

Weiss, A., Wortmeier, A.-K., & Kubicek, B. (2021). Cobots in Industry 4.0: A roadmap for future practice studies on human–robot collaboration. *IEEE Transactions on Human–Machine Systems*, *51*(4), 335–345. https://doi.org/10.1109/THMS.2021.3092684

Willingham, D. B. (1999). The neural basis of motor-skill learning. *Current Directions in Psychological Science*, *8*(6), 178–182. https://doi.org/10.1111/1467-8721.00042

Wong, S. W., Chan, R. H., & Mak, J. N. (2014). Spectral modulation of frontal EEG during motor skill acquisition: A mobile EEG study. *International Journal of Psychophysiology*, *91*(1), 16–21. https://doi.org/10.1016/j.ijpsycho.2013.09.004

Zhang, C., Zou, Y., Wang, F., del Rey Castillo, E., Dimyadi, J., & Chen, L. (2022). Towards fully automated unmanned aerial vehicle-enabled bridge inspection: Where are we at? *Construction and Building Materials*, *347*, 128543. https://doi.org/10.1016/j.conbuildmat.2022.128543

Zhao, Y., Chevrel, P., Claveau, F., & Mars, F. (2020, October). Towards a driver model to clarify cooperation between drivers and haptic guidance systems. In *2020 IEEE International Conference on Systems, Man, and Cybernetics (SMC)* (pp. 1731–1737). IEEE. https://doi.org/10.1109/SMC42975.2020.9283101

Zhu, F. F., Maxwell, J. P., Hu, Y., Zhang, Z. G., Lam, W., Poolton, J. M., & Masters, R. S. (2010). EEG activity during the verbal-cognitive stage of motor skill acquisition. *Biological Psychology*, *84*(2), 221–227. https://doi.org/10.1016/j.biopsycho.2010.01.015

About the authors

Fatemeh Dalilian is an Industrial and Systems Engineering doctoral candidate at the University of Iowa. Her research focuses on human decision-making and cognitive ergonomics in human-machine interaction, and she develops models and methods to support human cognition and behavior in complex systems.

David Nembhard is Professor of Industrial and Systems Engineering and Professor of Business Analytics at the University of Iowa. He researches workforce engineering and human cognition as it relates to human performance in complex systems, the science of learning and behavioral operations.