Emotional Turbulence During Simulation Training: Unraveling Emotion Dynamics and Performance Accuracy Using Simulations for Pilot Training

Alejandra Ruiz-Segura

Ph.D. in Educational Psychology

Department of Educational and Counselling Psychology

McGill University

April 2024

A thesis submitted to McGill University in partial fulfillment of the requirements of the Degree of Doctor of Philosophy in Educational Psychology

Table of Contents

Abstract

Affective responses, such as stress and emotions, are discussed as factors influencing human error in aviation; however, aviation research tends to explore affect as a subordinate element of cognition. Yet, more and more research is demonstrating that affective responses have a considerable effect on performance accuracy. Thus, the purpose of this dissertation is to examine the impact of pilots' affective responses on their flying performance accuracy in the context of training using simulations. The investigation started by synthesizing findings of previous research exploring affect and performance when using flight simulations, in a systematic literature review. Based on the findings of the literature review and previous research in education and psychology, two empirical studies were conducted aiming to understand the influence of emotions dynamics in flying performance. Emotions dynamics refer to the patterns and regularities characterizing fluctuations in emotions over time. Thus, the first study examines changes in performance and emotions dynamics (i.e., frequency, intensity, and variability of emotional responses) across training phases and difficulty levels in a flight simulation. The second empirical study builds on the previous one by examining the relationship between emotional variability and flying performance moderated by pilot trainees' perceived control and value over the task. The findings of this dissertation provide strong evidence that trainees' performance accuracy when performing simulated flying tasks is connected to their emotional responses. Particularly, findings from the empirical studies suggest that emotion dynamics might have adaptive functions for improving flying accuracy. Results demonstrate that affect and performance are dynamically impacted by training phases, difficulty levels, and subjective perceptions over the task. This dissertation concludes with a discussion of theoretical, methodological, and practical contributions, limitations, and future directions.

Résumé

Les réponses affectives, telles que le stress et les émotions, sont considérées comme des facteurs influençant les erreurs humaines dans le monde de l'aviation ; cependant, la recherche aéronautique tend à explorer l’affection comme un élément subordonné de la cognition. Pourtant, de plus en plus de recherches démontrent que les réponses affectives ont un effet considérable sur la performance. Ainsi, l’objectif de cette thèse est d’examiner l’impact des réponses affectives des pilotes sur la précision de leur vol dans le cadre d’entraînements avec des simulateurs. L'enquête a commencé par une revue systématique des recherches existantes qui explorent l'affection et la performance des pilotes lors de vols simulés À la lumière de ces résultats et de certaines théories de psychologie éducative, deux nouvelles études empiriques ont été menées dans le but de comprendre l'influence de la dynamique émotionnelle sur les performances de vol. Le concept de la dynamique émotionnelle représente les modèles et régularités des fluctuations émotionnelles au fil du temps. Ainsi, la première étude examine les changements dans les performances et dans la dynamique émotionnelle (c'est-à-dire la fréquence, l'intensité et la variabilité émotionnelles) lors des phases d'entraînement et lors des niveaux de difficulté dans une tâche de vol simulée. La deuxième étude empirique est la suite et examine la relation entre la variabilité émotionnelle et la performance de pilotes, modérée par les perceptions de contrôle et la valeur sur la tâche. Les résultats de cette thèse fournissent des preuves solides que la performance des pilotes lors des simulations est liée à leurs réponses émotionnelles. En particulier, les résultats suggèrent que la dynamique émotionnelle pourrait avoir des fonctions adaptatives pour améliorer la précision du vol. Les résultats démontrent que l'affection et la performance sont dynamiquement impactés par les phases d'entraînement, les niveaux de difficulté et les perceptions reliées aux tâches. Finalement, la conclusion inclut de nouvelles contributions théorique, méthodologique, et pratique à la littérature, les limitations de l’étude et finie sur quelques idées pavant la voie aux études futures.

Acknowledgements

I would like to thank my supervisor Dr. Susanne Lajoie for her professional and personal mentorship in my doctoral journey. First, I would like to thank Sue for accepting me as a student, and for enabling so many door-opening opportunities. I am very fortunate to have received Sue's academic guidance for developing multidisciplinary and innovative research, professional mentorship for developing project management skills, and for being a role model in social interactions, prioritizing people's individual differences.

I would like to thank my dissertation committee members, Dr. Jason Harley and Dr. Adam Dubé for their thoughtful advice in my comprehensive examination, research proposal, and this dissertation, their feedback helped improve this work. Thanks as well for sharing knowledge and advice to become a better researcher.

Thanks to previous members of Advanced Technologies for Learning in Authentic Settings for their academic, professional, and personal advice as well as current members for their company and feedback during this doctoral journey. Thanks to volunteers for their help in processing data and proactively improving processes.

Thanks for my family for their unconditional support in my adventures, and for your willingness to learn about my significant and unsignificant results. Thanks to my friends for being there in the ups and downs of the PhD and for believing in me. Special thanks to Frédéric Hanna, for his patient and loving company in the emotional ride of creating this dissertation.

I would like to acknowledge the project management support from all collaborators in the InLook Project (CAE Inc.; Marinvent Corporation; the Design Lab, Concordia University; the Faubert Lab, University of Montreal; Flight Research Laboratory, National Research Council) for their roles in experimental set-up and data collection, jointly funded by Natural Sciences and

Engineering Research Council of Canada (514052-17), Consortium for Aerospace Research and Innovation in Canada (CARIC) and Quebec (CRIAQ) (OPR-1618). This work was supported in part by the Fonds de recherche du Québec – Société et Culture Doctoral training scholarships (286648), and the Social Sciences and Humanities Research Council (SSHRC) Grant of Canada awarded to Prof. Susanne Lajoie and the Learning Across the Disciplines and Canada Research Chair funding on Advanced Technologies in Authentic Settings 950-231835. Special thanks are addressed to Hugh Grenier and Alain Bourgon (CAE) as the subject-matter experts for designing the flying exercises, training researchers, and creating the video training for participants, and David Bowes, Alexander Ireland, and Jack Russ (CAE) for the management of this project. Special thanks to Joao Araujo and Dr. Sujoy Ghosh Hajra for their important contributions to system integration and experimental data collection.

Compliance with Ethical Standards

Conflict of Interest. The co-authors that participated in this dissertation declare that they have no known conflict of interest.

Ethical Approval. All procedures performed in human participants were in accordance with the ethical standards from Concordia University Research Ethics Board and was accepted by McGill University based on the CREPUQ interinstitutional agreement, in the project titled "Inferring pilot trainee's affective and cognitive states from biometric data during the training processes".

Informed Consent. Informed consent was obtained from all individual participants included in the study.

Dedication

To my mom, who taught me the value of education; to my dad, who always validated my emotions; and to my sister, who reminds me to stay focused.

Contributions to Original Knowledge

This dissertation makes multiple contributions to knowledge including theoretical, methodological, and practical inputs. Concretely, this manuscript-based dissertation contributes to knowledge by creating and publishing three original scientific articles. This section shows brief contributions to knowledge of each manuscript; however, detailed contributions are presented in corresponding chapters. The general theoretical, methodological, and practical implications are discussed in detail in Chapter 5, final discussion.

Manuscript 1 (Chapter 2)

This manuscript is a systematic review that contributes to understanding the relationship between affect (emotions and stress) and performance when training with flight simulations. To our knowledge this is the first systematic review of this kind.

Ruiz-Segura, A. & Lajoie, S.P. (in press). Affect and performance in simulated flying tasks: A systematic literature review. *The International Journal of Aerospace Psychology.*

Manuscript 2 (Chapter 3)

This manuscript contributes to understanding emotion dynamics in flight training by exploring the frequency, intensity, and variability of emotions across training phases and difficulty levels, which, to our knowledge, has not been studied before.

Ruiz-Segura, A. Law, A., Jennings, S., Bourgon, A., Churchill, E., & Lajoie, S. (under review). Flight emotions unleashed: Navigating training phases and difficulty levels in simulated flying. *The Journal of Computer Assisted Learning.*

Manuscript 3 (Chapter 4)

This empirical study contributes to our understanding of the moderating role of perceived control and value in explaining the association of emotional variability and flying performance, in the context of training beginner pilots using simulators.

Ruiz-Segura, A., Law, A., Jennings, S., Bourgon, A., Churchill, E., & Lajoie, S. (under review). A study of pilot trainees' perceptions, emotional variability, and performance. *Aerospace Medicine and Human Performance*.

Contribution of Authors

I am the primary author on all the papers included in this dissertation and I am responsible for their content. I wrote each chapter independently. Dr. Susanne Lajoie provided feedback for each chapter and revised this dissertation in full. Chapter 2 was inspired from my comprehensive exams for which Dr. Adam Dube, Dr. Jason Harley provided feedback; I repurposed Dr. Dube’s and Dr. Harley’s comments for writing Chapter 2, co-authored with Dr. Susanne Lajoie and accepted with minor revisions to the Journal of Aerospace Psychology. The manuscripts in Chapter 3 and Chapter 4 are co-authored with Dr. Andrew Law, Sion Jennings, Alain Bourgon, and Ethan Churchill, who provided feedback for improving the manuscript before submitting to journals. Specifically, for both Chapter 3 and Chapter 4, I independently conceptualized and wrote the manuscripts in full. Dr. Law and Jennings contributed with software analysis and the methodology for analyzing flying performance data. Ethan Churchill contributed to data curation and visualization. Dr. Law, Jennings, Bourgon, and Dr. Lajoie contributed to editing and reviewing the manuscripts in full. In Chapter 3, Dr. Law, and Dr. Lajoie provided input on data analysis techniques. Chapter 3 was submitted, and revisions were invited to the Journal of Computer Assisted Learning and Chapter 4 was submitted and revisions were invited to the Aerospace Medicine and Human Performance Journal. The conclusions and contributions in each chapter are original, and they provide contributions to the advancement of knowledge.

List of Tables

List of Figures

Chapter 1. Introduction

During 2024, it is predicted that there will be a gap of 15% between pilots' supply and demand, and if the aviation industry does not implement changes, this gap will continue to grow (Murray et al., 2022). To face the growing flight demands, it is necessary to increase the pilot supply, particularly by training skilled pilots who can perform accurately, even in complex situations.

Simulations have been a basic tool for flight training since World War II (Hamman, 2004), helping beginner pilot trainees familiarize themselves with the management of the cockpit. In this dissertation, simulations are used as an umbrella term for techniques that are used for mimicking and replacing real environments in which learners can deliberately practice their skills (Gaba, 2004; Lajoie, 2021). Following this definition, this dissertation focuses on technology-based simulations, ranging between computer-based simulations, high-fidelity simulators constructed copying an aircraft cockpit, and in-flight simulations that imply the use real aircrafts to perform simulated tasks (Lajoie, 2021; Weingarten, 2005). Specifically, we explored beginner pilot trainees’ learning process with a simulated flying task.

Research exploring learning and training has demonstrated that performance accuracy is influenced by learners' cognitive and affective processes while they are solving the learning task (D’Mello & Graesser, 2012; Pekrun, 2019). Thus, pilot trainees' accuracy is likely influenced by the affective processes they are experiencing during flight training. However, models of aviation explaining flight performance only account for affective responses as subordinate to cognitive processes. Cognitive processes are hypothesized to have a direct link with performance accuracy based on attentional control theory and the situational awareness theory (Endsley, 2000; Eysenck et al., 2007). Yet, there are educational theories that describe a more direct role of affect in learning. In this dissertation it is argued that affective responses will directly account for pilot trainees' performance accuracy.

The purpose of this dissertation is to bridge the gap between educational and aviation research by exploring the role of discrete emotions and emotion dynamics and their relation to performance accuracy in simulated flying tasks. The studies included in this dissertation attempt to answer the questions: 1. How do emotions and performance dynamically change during flight simulation training?; and 2. How do emotions and performance relate to each other during simulated flying tasks?

According to our findings, emotion dynamics have not been explored in depth in learning contexts (Kuppens, 2015; Zheng et al., 2023a). Emotion dynamics refer to the patterns and regularities of emotional shifts and fluctuations in a given period of time (Bailen et al., 2019; Krone et al., 2018; Kuppens, 2015; Zheng et al., 2023a). Emotions in professional training contexts have been explored mostly as frequency of discrete emotions (like frustration, joy, surprise, etc.) (e.g., Artino et al., 2010; Landman et al., 2020; Rosa et al., 2022). However, recent approaches are suggesting that studying emotional dynamics in training contexts might provide a wider and more detailed perspective of the role of emotions in training (Hou et al., 2023; Zheng et al., 2023a). Thus, this dissertation will explore a combination of discrete emotional expressions and dynamic patterns to understand the role of affect in flying performance accuracy.

Overview of the Chapters

Chapter 2 presents a systematic literature review exploring the role of affect in flight training using simulations. This systematic literature review focused on affective changes that can occur within a commercial flight, regulated to last a maximum of 16 hours in Western countries (Canadian Aviation Regulations, 2022; Code of Federal Regulations, 2023; European

Commission, 2013). For that reason, we concentrate on stress and emotions that last from seconds to minutes, and minutes to hours accordingly (Gross, 2015). Stress is considered to be inextricably linked to negative-activating emotions of fear and anxiety (Fink, 2016); however, fear and anxiety can be seen as points along a continuum, whereas stress is longer lasting and less event specific, thus containing expressions of discrete negative-activating emotions (Craske et al., 2009). Although moods may impact affect and flying performance, their longer duration (i.e., days to weeks) might be less specific to flying due to their long durations and because moods are less object-specific, being referred to as the "affective weather", thus making it more complex to identify if those affective states are triggered by the flight itself (Gross, 2015; Jarrell & Lajoie, 2017). The empirical articles included were categorized in themes according to the manner in they contribute to answering the following question: What is the relation between affect and flying performance when training with flight simulations? The findings of this literature review identified current trends in this field and show areas that need further exploration. Thus, the results of this systematic review guided the creation of the empirical studies, presented in Chapter 3 and 4, intended to move the field forward.

An insight obtained from the Chapter 2 is that some previous studies exploring affect and performance lacked a consistent theoretical basis, resulting in inconsistent interpretation of results. Therefore, the empirical Chapters 3 and 4 are guided by Pekrun's control-value theory of achievement emotions (2019). This theory assumes that emotions in activities that lead to success or failure, i.e., achievement situations, are triggered by learners' sense of agency and importance of the task, namely by appraisals of control and value. The control-value theory follows the assumption that emotions unfold in context, thus, the characteristics of the task, and the domain where the activity takes place will help determine the emotions to be expected

(Pekrun, 2019). Lastly, the type of emotions that learners experience is expected to have an impact on the performance outcome. Generally, positive-activating emotions, like joy, tend to be related to more successful performance, whereas negative-deactivating emotions, like boredom, are related to less desirable outcomes. This dissertation is theoretically guided by the control-value theory, yet it is expected that the empirical chapters will contribute to understand the applications of this theory in the context of flight training with simulated tasks.

Chapter 3 is an empirical study that attempts to contribute to the understanding of beginner pilot trainees' performance and emotional experience across training phases and difficulty levels, using a multimodal approach. Performance accuracy was evaluated by the distance between the target metric and the aircraft position, using the simulator logfiles and an expert evaluation of performance. This study explores behavioral and physiological expressions of emotional intensity: facial expression and electrodermal activity. This study contributes to understanding emotion dynamics by exploring emotional variability, i.e., fluctuations across multiple emotions.

Chapter 4 attempts to bridge the gap between educational and aviation research by exploring cognitive appraisals of emotions in a simulated flying task. Previous studies gathered in Chapter 2 demonstrate that pilots perceptions of the task help explain the relationship between stress and flying performance (Hart & Bortolussi, 1984; Vine et al., 2015). The control-value theory explains that learners' appraisals of control and value are antecedents of achievement emotions. Therefore, appraisals can help us understand the conditions under which emotions and performance relate. However, to our knowledge, the relationship between control and value appraisals and emotional variability has not been explored in the context of flight training. Additionally, a limitation of Chapter 3 was that it lacked information regarding trainees' subjective experience of the task. Therefore, Chapter 4 explores the moderating role of trainees' subjective perception of control and value over the task to explain the relationship between emotional variability and flying performance.

Lastly, Chapter 5 concludes this dissertation by interpreting the findings and specifying the theoretical and methodological contributions, as well as the limitations of this research. This chapter presents possibilities for future research directions in this field, along with suggestions that can lead to improvements in practice.

References

Artino, A. R., Rochelle, J. S. L., & Durning, S. J. (2010). Second-year medical students' motivational beliefs, emotions, and achievement. *Medical Education*, *44*(12), 1203–1212. https://doi.org/10.1111/j.1365-2923.2010.03712.x

Bailen, N. H., Green, L. M., & Thompson, R. J. (2019). Understanding emotion in adolescents: A review of emotional frequency, intensity, instability, and clarity. *Emotion Review*, *11*(1), 63–73. https://doi.org/10.1177/1754073918768878

Canadian Aviation Regulations. (2022, December 21). *Commercial air services*. Justice laws website - Government of Canada. https://laws-lois.justice.gc.ca/eng/regulations/sor-96-433/page-83.html

Code of Federal Regulations. (2023, April 19). *14 CFR 91.1059 -- Flight time limitations and rest requirements: One or two pilot crews.* Electronic code of federal regulations. https://www.ecfr.gov/current/title-14/chapter-I/subchapter-F/part-91/subpart-K/subject-group-ECFRc17623c0e0be17e/section-91.1059

Department of Defense. (2011). *Modeling and simulation glossary*. Modeling and Simulation Coordination Office. https://www.acqnotes.com/Attachments/DoD%20M&S%20Glossary%201%20Oct%2011.pdf

D'Mello, S., & Graesser, A. (2012). Dynamics of affective states during complex learning. *Learning and Instruction*, *22*(2), 145–157. https://doi.org/10.1016/j.learninstruc.2011.10.001

Endsley, M. R., (2000). Theoretical underpinnings of situation awareness: A critical review. In M. R. Endsley & D. J. Garland (Eds.), *Situation Awareness Analysis and Measurement*. (pp. 3-28). Lawrence Erlbaum Associates, Inc.

European Commission. (2013, October 4). *Pilot and crew fatigue—Frequently asked questions*. Directorate-General for Communication, European Commission. https://ec.europa.eu/commission/presscorner/detail/en/MEMO_13_854

Eysenck, M. W., Derakshan, N., Santos, R., & Calvo, M. G. (2007). Anxiety and cognitive performance: Attentional control theory. *Emotion*, *7*(2), 336–353. https://doi.org/10.1037/1528-3542.7.2.336

Gaba, D. M. (2004). The future vision of simulation in health care. *Quality & Safety in Health Care*, *13 Suppl 1*(Suppl 1), i2-10. https://doi.org/10.1136/qhc.13.suppl_1.i2

Gross, J. J. (2015). Emotion regulation: Current status and future prospects. *Psychological Inquiry*, *26*(1), 1–26. https://doi.org/10.1080/1047840X.2014.940781

Hamman, W. R. (2004). Will simulation fly in medicine as it has in aviation? *Quality & Safety in Health Care*, *13*(5), 397–399. https://doi.org/10.1136/qhc.13.5.397

Hart, S. G., & Bortolussi, M. R. (1984). Pilot errors as a source of workload. *Special Issue: Aviation Psychology*, *26*(5), 545–556.

Hou, C., Zhu, G., & Yang, Y. (2023, June 10). *Beyond the cognitive dimension: Emotion patterns in productive and improvable knowledge building discourse*. Annual Meeting of the International Society of the Learning Sciences. https://doi.org/10.22318/cscl2023.333889

Krone, T., Albers, C. J., Kuppens, P., & Timmerman, M. E. (2018). A multivariate statistical model for emotion dynamics. *Emotion*, *18*(5), 739–754. https://doi.org/10.1037/emo0000384

Kuppens, P. (2015). It's about time: A special section on affect dynamics. *Emotion Review*, *7*(4), 297–300. https://doi.org/10.1177/1754073915590947

Lajoie, S. P. (2021). Multimedia learning with simulations. In L. Fiorella & R. E. Mayer (Eds.), *The Cambridge handbook of multimedia learning* (3rd ed., pp. 461–471). Cambridge University Press. https://doi.org/10.1017/9781108894333.048

Landman, A., van Middelaar, S. H., Groen, E. L., van Paassen, M. M., Bronkhorst, A. W., & Mulder, M. (2020). The effectiveness of a mnemonic-type startle and surprise management procedure for pilots. *The International Journal of Aerospace Psychology*, *30*(3–4), 104–118. https://doi.org/10.1080/24721840.2020.1763798

Murray, G., Heilakka, R., & Medland, A. (2022, November 1). *The pilot shortage will only get worse as the decade progresses*. BRINK – Conversations and Insights on Global Business. https://www.brinknews.com/the-pilot-shortage-will-only-get-worse-as-the-decade-progresses/

Pekrun, R. (2019). Achievement emotions: A control-value theory perspective. In Patulny, A. Bellocchi, R. Olson, S. Khorana, J. McKenzie, & M. Peterie (Eds.), *Emotions in Late Modernity* (pp. 142-157). Routledge.

Rosa, E., Lyskov, E., Gronkvist, M., Kolegard, R., Dahlstrom, N., Knez, I., Ljung, R., & Willander, J. (2022). Cognitive performance, fatigue, emotional, and physiological strains in simulated long-duration flight missions. *Military Psychology*, *34*(2), 224–236. https://doi.org/10.1080/08995605.2021.1989236

Vine, S. J., Uiga, L., Lavric, A., Moore, L. J., Tsaneva-Atanasova, K., & Wilson, M. R. (2015). Individual reactions to stress predict performance during a critical aviation incident. *Anxiety, Stress & Coping: An International Journal*, *28*(4), 467–477.

Weingarten, N. C. (2005). History of in-flight simulation at general dynamics. *Journal of Aircraft*, *42*(2), 290–298. https://doi.org/10.2514/1.4663

Zheng, J., Li, S., & Lajoie, S. P. (2023). A review of measurements and techniques to study emotion dynamics in learning. In V. Kovanovic, R. Azevedo, D. C. Gibson, & D. lfenthaler (Eds.), *Unobtrusive observations of learning in digital environments* (pp. 7–29). Springer. https://doi.org/10.1007/978-3-031-30992-2_2

Chapter 2. Manuscript 1

Affect and Performance in Simulated Flying Tasks: A Systematic Literature Review

Ruiz-Segura, A. & Lajoie, S.P. (in press). Affect and performance in simulated flying tasks: A systematic literature review. *The International Journal of Aerospace Psychology*

Abstract

Objective. This systematic review explores the relationship between affect and performance during flight simulations in the context of training using simulated flying tasks.

Background. A goal of human factors is to understand pilots' psychological processing to reduce human error. Psychological insights have mainly focused on cognition, whereas recent approaches suggest that affective responses (stress and emotions) can impact performance accuracy. This is the first literature review exploring affect and performance in simulated flying tasks.

Method. Following the PRISMA framework, a systematic review was conducted, resulting in 29 articles meeting the inclusion criteria.

Results. Studies were grouped in themes according to the manner they contributed to our objective. Findings revealed that: 1) affect and flying performance are continuous processes that need to use continuous and non-invasive measurements; 2) negative-activating affect (stress and anxiety) tend to be detrimental for flight performance; 3) pilot-co-pilot and pilot-instructor interactions induce affect in simulated flight; 4) unusual flying scenarios induce negative and activating affective responses and decrease performance accuracy; 5) interventions are successful for pilots to manage performance and affect.

Conclusion. This systematic review demonstrates the relationships between affect and performance. The main connection is that unpredicted and intense affective reactions can distract pilots, resulting in errors. Simulations can induce authentic affective reactions, allowing pilot trainees to familiarize themselves with their reactions, and regulate themselves, to achieve successful performance.

Keywords: Flight training, emotions, stress, simulation, flying performance.

Introduction

Human factors in aviation cover "psychological, physical, biological, and safety characteristics of individuals and groups" (Helmreich, 2010, pp. ix). While cognitive and social psychological processes, like workload, fatigue, and communication, have been extensively studied in human factors, the emotional component is often overlooked or considered secondary to cognition (Eysenck et al., 2007; Hart, 2006; Salas et al., 2010). Research in educational psychology highlights the intertwined nature of affective responses and cognitive processes impacting performance (D'Mello & Graesser, 2012; Pekrun, 2019).

This systematic review contributes to the psychological aspect of human factors in aviation by focusing on the affective component. Specifically, we investigate the relationship between affect and performance during flight simulations, particularly training. Given the rapid technological advances in aviation and the demand for pilots to adapt to these changes (Salas et al., 2010), we emphasize technology-based simulations.

Flight Simulations for Training

Simulations are "technique[s], not technolog[ies], to replace or amplify real experiences with guided experiences […] replicat[ing] substantial aspects of the real world in a fully interactive manner" (Gaba, 2004, pp. i2). Simulations offer safe and authentic environments for pilots deliberate practice of technical skills and better understand their affective reactions (Gordon et al., 2010; Lajoie, 2017). Simulations are crucial in high-stakes scenarios, enabling trainees to experience intense affective reactions in a controlled manner (Morris et al., 2004). Immersive simulations help trainees familiarize themselves with their behavioural and affective reactions, enhancing their ability to manage their cognitive resources efficiently to reach successful performance (Cross et al., 2023; Pekrun, 2019).

Merging aviation with education and training, simulations can be seen as technology-rich learning environments designed to support, extend, and transform instructional situations (Lajoie & Poitras, 2023). We include technology-rich learning simulations, namely high-fidelity simulators, in-flight simulations, computer-based instruction, intelligent-tutoring systems, virtual, augmented, or mixed reality, and serious games (Lajoie et al., 2020; Lajoie & Azevedo, 2006; Salas et al., 2010; Weingarten, 2005).

We define *simulations* as a technique and *simulators* as the high-fidelity technological devices that physically mimic aircrafts (Gaba, 2004). Computer-based simulations are platforms to support knowledge and skill acquisition through system embedded tools (Harley, 2016; Jarrell et al., 2017). They differ from simulators and in-flight simulations as they are complementary learning tools, whereas simulators and in-flight simulations are supplements to risky real-life situations (Roh & Jang, 2017). In-flight simulations imply the use of an aircraft; however, their objective is to develop research and training; and rather than using computer-based responses, they maintain the real motion and visual cues (Weingarten, 2005). Virtual reality (VR) facilitates sensorial immersion to reproduce physical or behavioural events, giving users the experience of being in a simulated reality (Makransky et al., 2019; Parong & Mayer, 2020). Augmented reality (AR) amplifies real-world experiences by adding computer-generated elements and multimedia contents (Hodhod et al., 2014; Martín-Gutiérrez et al., 2017). Serious games are videogames for learners to practice skills, rather than entertainment (Muñoz et al., 2022; Pilote & Chiniara, 2019).

Flying Performance

Human factors research aims to inform opportunities to improve performance and reduce human error (Bent & Chan, 2010). Performance can be evaluated according to an end goal (accuracy), or accounting for the processes taken to accomplish a task (efficiency) (Jarrell et al., 2017). Accuracy refers to task correctness (Li & Lajoie, 2021), while efficiency measures the deviation from ideal solution (McClernon & Miller, 2011). We explore the impact of affective responses on accuracy and efficiency in performance using flight simulations (D'Mello & Graesser, 2012; Pekrun, 2019). Cognition, affect, and performance are distinct but interconnected. For instance, pilots' decision making, and intense emotions will influence their performance accuracy (Martins, 2016). This manuscript emphasizes affect's role in flight performance.

Affect while Flying

Affect, along with cognition and conation, is a core component of the mind (American Psychological Association, n.d.). Affect includes internal states like stress, emotions, and moods, quickly evaluating what is "good or bad for me" (Gross, 2015). Affect responses are categorized by valence and arousal (Russell, 2003). Valence indicates subjective experience of a positive or negative feelings, while arousal refers to changes in physiological activation (Russell, 2003).

The three components of affect (i.e., stress, mood, and emotion) differ in their duration, expression, and function (Gross, 2015). Emotions are brief states, lasting from seconds to minutes, that function as an automatic response to evaluate specific stimuli. Stress is an arousing reaction during challenging situations, the trigger is unspecific, lasting from minutes to hours. Mood is identified, the affective "weather", is a fuzzy state that lasts from days to weeks (Gross, 2015; Jarell & Lajoie, 2017). This manuscript focuses on affective within a commercial flight, having a maximum duration of 14 hours (Code of Federal Regulations, 2023; European Commission, 2013). We exclude moods for their long-term duration, as the effect on flight performance might be less direct.

All affective responses have a function, aiding instinctive decision-making actions (Damasio, 2005). This manuscript explains the function of affect following the Yerkes-Dodson law (1908) and the control-value theory of achievement emotions (CVT) (Pekrun, 2019). Stress, a response to the inability to meet situational demands, prompts individuals to seek resources for resolution (DeMaria et al., 2010; Landman et al., 2017). Stress entails subjective, behavioural, and physiological changes. We recognize that there are situations where eustress (or good stress) can be beneficial for learning (Rudland et al., 2020). The Yerkes-Dodson law (1908) explain the relationship between stress and performance visualized in an inverted U-shape, very low or very high levels of stress would be detrimental to performance; a stimulating, but not overwhelming degree of stress can stimulate pilots to seek solutions.

Emotions are also multi-componential affective states, they have a specific object-focus, shorter duration, and can be identified in a valence-arousal continuum (Gross, 2015; Pekrun, 2019). CVT explains the relationship between emotions and performance. Positive-activating emotions like joy, are more functional since their agreeable experience motivate learners to re-take the task, and the physiological activation stimulate learners for reaching the learning objective (Pekrun & Perry, 2014). Negative-deactivating emotions, like boredom, tend to be less functional, since the learners are demotivated by disagreeableness, and not physiologically aroused to act (Goetz & Hall, 2014). Positive-deactivating emotions, like relaxation, might motivate learners to engage in similar tasks in the future, yet in the short term, the physiological deactivation may be unhelpful as learners may not feel the need to try hard (Pekrun & Perry, 2014). Negative-activating emotions, like anxiety, have the opposite effect: the physiological arousal might activate learners seek solutions, but the displeasing feeling might lower their motivation to the future (Pekrun & Perry, 2014). Since both affect and performance are context based, we identify functional affective reactions for achieving desired performance in flight training.

This systematic review explores affective changes within commercial flights (maximum 16 hours). Previously, Li (2020) conducted a literature review exploring biometric measures used to understand affect in flight training contexts. Li's review (2020) did not explore associations with flying performance. Furthermore, Martin et al. (2015) conducted a review on startle in aviation. Startle is the automatic reaction that occurs as a defense of alarming situations, closely related to surprise; however, they did not explore other emotions. Neither Li (2020) nor Martin et al. (2015) examined the role of affect using simulations.

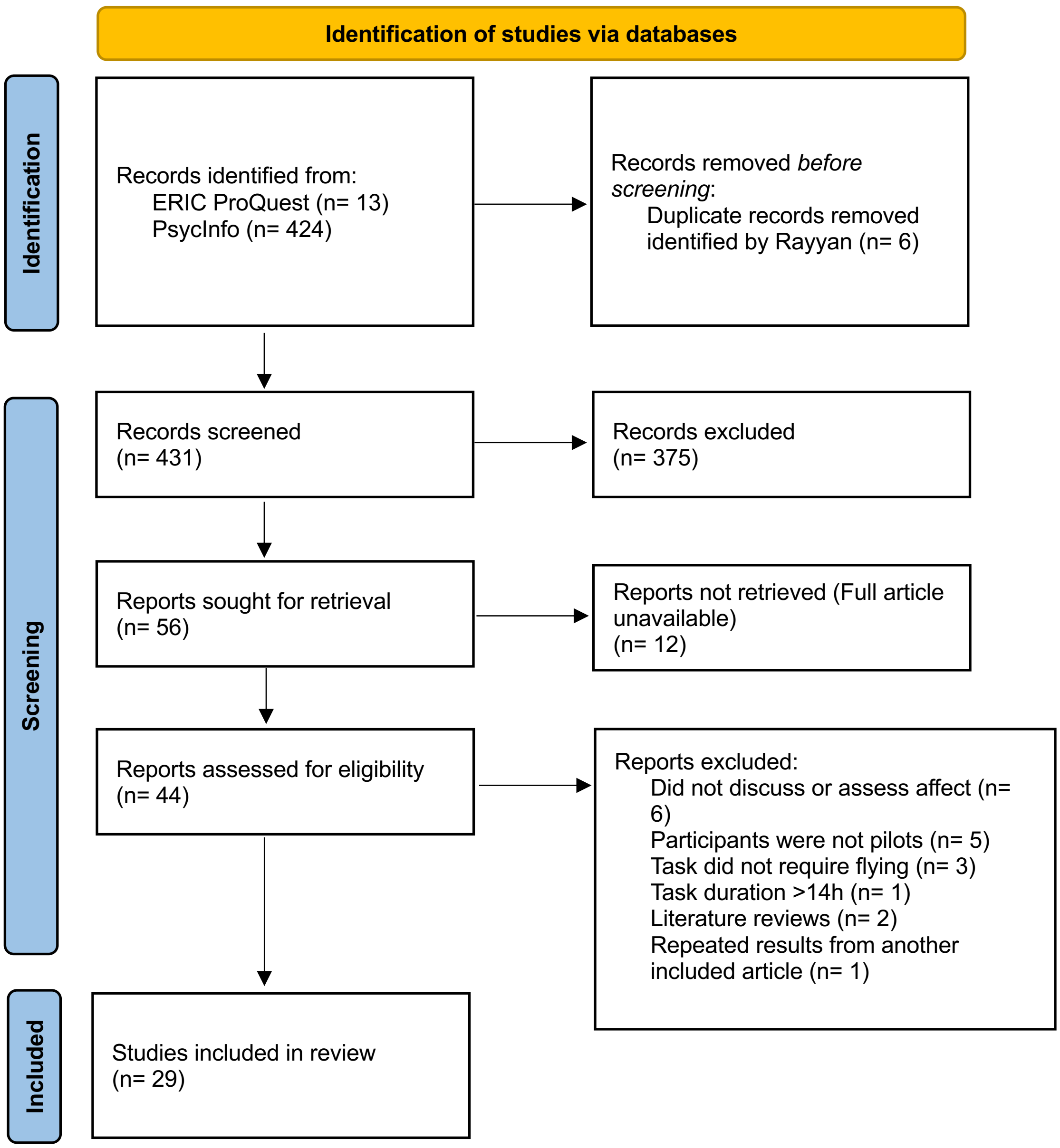
This systematic review fills a gap by exploring the relationship between affect and performance in simulated flying tasks, investigating simulations' authenticity in documenting affect and its impact on flying performance. The research question is: What is the relation between affect and performance when training with flight simulations?

Methods

A systematic review was conducted to examine empirical articles exploring the connections between affect and performance when using flying simulations. We followed the PRISMA framework to guide the selection process, attempting to make it trustworthy and replicable (Moher et al., 2009; Page et al., 2021). Figure 1 shows the search strategy.

Figure 1

Systematic Review PRISMA Flow Chart



We selected two scientific databases in education and psychology domains, ERIC ProQuest and PsycInfo. To broaden our search, we added a general scientific database, SCOPUS. We used the strings on Table 1 combined with the Boolean operator “AND”. We followed inclusion and exclusion criteria to guide the screening process, see Table 2. In the identification stage we used the database operator to include only peer-reviewed articles.

Table 1

Search Strings Used in Database Search

Concept	Search string
Pilots	pilot trainee OR aviation trainee OR flyer OR copilot OR aircrew OR aviator OR aviatrix OR captain OR commander OR flight OR flight trainee OR student pilot
Affect	affect* OR emotion* OR stress OR positive emotion* OR negative emotion* OR mental stress OR acute stress OR psychological stress OR distress
Simulation (as technology-rich learning environment)	technology rich learning environment* OR advanced learning technolog* OR multimedia learning environment OR technology enhanced learning OR virtual reality OR gamification OR serious game* OR augmented reality OR virtual reality OR multimedia learning OR intelligent tutor* OR interactive technology OR computer-based learning environment OR simulat*

The searches resulted in 441 articles. SCOPUS suggested five books that were excluded. Hence, SCOPUS is not included in the PRISMA chart. For the remining 436 articles, we used Rayyan, a specialized application to screen abstracts for systematic reviews (Ouzzani et al., 2016). Using Rayyan's algorithms, six duplicates were deleted. We screened the title and abstract of 430 articles, resulting in 374 excluded articles. We conducted a full-text review of the 56 included articles. 28 articles were excluded: 12 articles were not available, six did not discuss or assess affect, five did not include or mimic pilots, three did not require flying (i.e., unmanned aerial vehicles), two were literature reviews, and one repeated identical results. We included an article that met the criteria but was not found through the systematic search (i.e., Cao et al., 2019).

Table 2

Search Inclusion and Exclusion Criteria

Inclusion	Exclusion
• Participants are professionals, trainees, or students receiving flight training. • Discuss or assess emotion and stress implications. • Discuss or assess flying performance. • Uses simulations. • Peer-reviewed	• Participants are professionals not flying aircrafts (i.e., flight assistants, astronauts). • Discuss or assess mood, chronic stress, or mental diseases. • Used aviation tasks that do not require pilots flying (i.e., unmanned aerial vehicles, air battle management) • Non-peer reviewed • Full books • PowerPoint presentations or abstract-only publications • Inaccessible full text. • Experiments conducted in animals or children.

Our systematic review resulted in 29 articles. Key sections were synthesized during the full-text review, including research objectives, participants, simulation type, affect measurement, performance measurement, and findings.

Findings

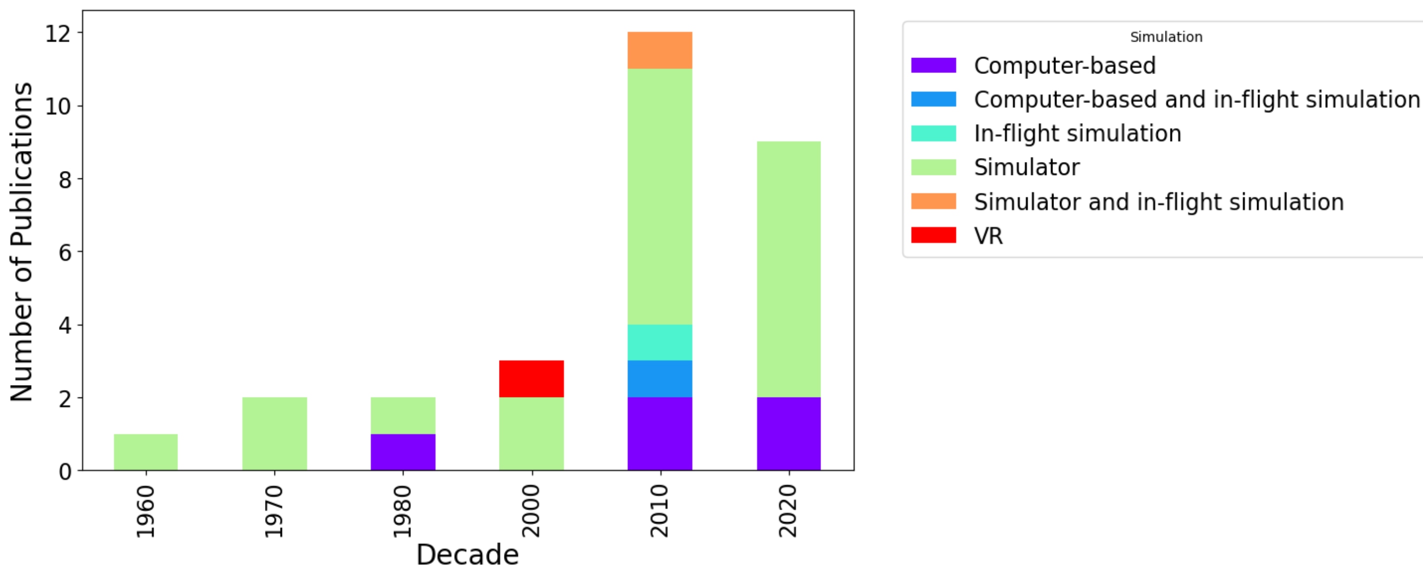
The selected studies were published between 1968 and 2024, with 86% published after the year 2000 (n=25). Sample sizes ranged from six to 90 participants (M=27.69, SD=19.48). 75.86% (n=22) included participants with flying experience, seven studies included participants with no prior flying experience. We will refer to "participants" to identify non-pilots, "student pilots" to refer to pilots still training, and "pilots" will be used for licensed and experienced pilots.

Simulation types varied, with simulators being predominant in 20 of the 29 studies. Computer-based simulations are the second most used (n=5), while virtual reality (VR), in-flight simulations, combination of simulator and in-flight simulation, and computer-based simulation and an in-flight simulation were less frequent. The most studied affect concepts are emotions (*n*=12) and stress (*n*=11), and a combination of both (*n*=3). The term "affect" (*n*=3) was used less frequently. Figure 2 shows the distribution of publication per decade. 2-A depicts simulation type, and Figure 2-B illustrate the affect concepts used.

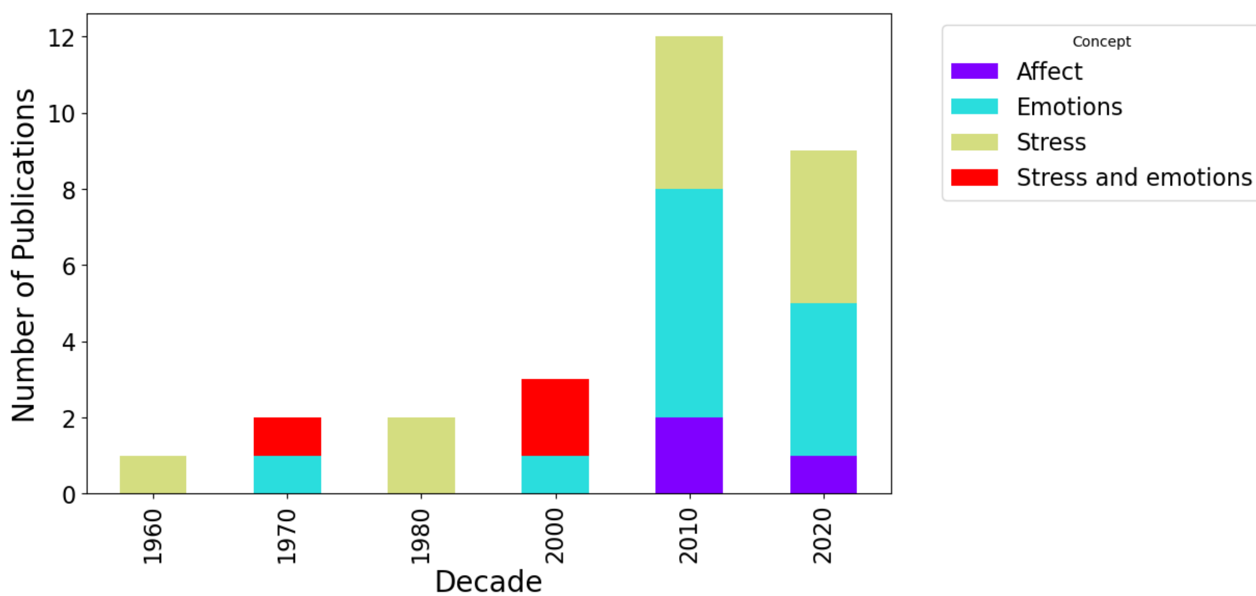
Figure 2

Distribution of Publications per Decade

A) Simulation type



B) Affect term



Connections Between Affect and Flying Performance

Five themes emerged during the synthesis process regarding the research question: What is the relation between affect and performance when training with flight simulations? Themes were categorized by their approach to explain affect and performance. Studies in Theme 1 explore affect and flying performance, but do not empirically analyse their relationship. Conversely, themes 2 and 4 explore the relationship between flying performance and specific affective responses descriptively (negative-activating affect and surprise, accordingly). Theme 3 focuses on how social interactions influence affect and flying performance, rather than exploring a specific affective response. Theme 5 includes empirical studies developing interventions to modify affect and flying performance.

Theme 1. Affect And Flying Performance Are Continuous Processes

Studies in Theme 1 explore continuous, non-invasive measures to gauge pilot performance and affect without distracting them during simulations (Mathavara et al., 2022). A pioneer study measured pilots' stress (n=18) using jaw muscle tension and blinking rate measures (Drinkwater et al., 1968). Stress was induced by electrical shocks on pilots' feet when having errors, correlating with higher muscle tension and higher blinking rate. Higher muscle tension related to less accuracy in reaching the designed navigational path.

Recent studies explored affect triggered by flying tasks. Silva et al. (2009) tested a model for detecting valence and arousal automatically when using VR. Flight characteristics were manipulated to induce affect according to Russell's circumplex model (2003). Weather conditions were hypothesized to associate with valence, with fair weather triggering positive emotions, and bad weather causing negative ones. Flight routes would change arousal, low arousal had open, easier, turns, whereas high arousal was caused by routes with close, more complicated, turns. Valence was inferred from electrodermal activity (EDA). EDA is known to vary with arousal, interpreted following the Yerkes-Dodson law where too low or too high skin conductivity associates to poorer performance (Braithwaite et al., 2015); the study lacked a rationale for using EDA to infer valence. Arousal was inferred from temperature and respiration rate. After performing the tasks, participants (n=37) pointed to their experienced emotions in a visual representation of Russell's circumplex model. Participants' descriptions aligned 77% with physiological valence and arousal. This innovative study uses continuous, non-invasive measures to infer valence and arousal; however, it lacked details guiding the selection of measures, limiting interpretation and reproduction.

Guided by Pekrun's (2019) theory, a study explored participants' (n=19) changes in valence and arousal as they performed basic maneuvers in computer-based simulations (Li & Lajoie, 2021). Emotions were labelled using facial expression, used to infer behavioural valence and arousal. EDA was used to infer physiological arousal, and questionnaires were used to understand antecedents of emotions (Pekrun, 2019). Results showed that behavioural and physiological arousal increased together. Perceived task value negatively related to behavioural arousal, whereas perceived control was negatively related to expressing fear. Physiological arousal positively predicted performance in low difficulty tasks but had no significant correlations in medium and high difficulty tasks, nor among participants accuracy. This study provides an example of theoretically based measurements facilitating interpretation of psychological constructs.

Other studies explored emotional and cognitive demands during long flight simulations (12 hours), using questionnaires for labelling emotions, and heart rate for identifying physiological changes (Rosa et al., 2021, 2022). Participants (*n*=12) self-reported emotional changes at baseline, middle, and end of the task (Rosa et al., 2021). Overtime, participants were less cheerful, stimulated, and enthusiastic, and more bored. Poorer cognitive performance (i.e., increase in response time) was negatively correlated to positive emotions and positively correlated to negative ones. Self-reported emotions and heart rate variability (HRV) features correlated during the flight. HRV indicates the variability in the length of time between heart beats, thus lower HRV tends to associate to higher negative-activating affect (Cao et al., 2019). Increases in HRV positively associated to sadness, boredom, and inactivity, and negatively associated to feeling cheerful, enthusiastic, and stimulated. Authors explained this as a decrease in energy over time. This study suggests that affect measurements should be triangulated with a grounded truth measure visualizing trainees' perspective (Harley, 2016).

More studies explored pilot trainees' emotional profiles using continuous measures of emotions. Tichon et al. (2014) used pupillometry to infer student pilots (*n*=12) affective changes according to difficulty levels in a flight simulator. Participants' eye-saccades and saccade velocity negatively related to self-reported positive affect. More anxious participants had more saccades and pupil size increased, whereas less anxious participants had constant saccades and pupil size decreased during difficult tasks. Similarly, Gaetan et al. (2015) explored emotional patterns of pilots (*n*=6) flying helicopter simulators. Emotions were measured with muscle tension, EDA, and questionnaires. Three profiles were described. Expert-like pilots' physiological arousal increased with difficulty level and reported experiencing positive emotions with low intensity. Non-experts' physiological arousal increased with perceived workload, i.e., the perceived relationship between mental resources and task demands (Hart & Staveland, 1988), and they experienced positive and negative emotions with high variability and high intensity.

Intermediate pilots reported positive emotions, but their physiological arousal increased with cognitive workload (rather than difficulty level). These studies reveal potential classifications to understand pilots' emotional profiles according to affective patterns and expertise.

Human error is linked to pilots' cognitive flaws (Mathavara et al., 2022). Performance requires continuous high concentration, where distractions can lead to error. Consequently, affect must be considered and assessed in a continuous and non-distracting manner. Studies in theme 1 demonstrate that affect can be inferred in non-invasive manners from muscle tension, pupillometry, electrodermal activity, facial expression, self-reported measures, and heart rate variability, reducing pilot trainees cognitive distractions (Drinkwater et al., 1968; Gaetan et al., 2015; Li & Lajoie, 2021; Rosa et al., 2021, 2022; Silva et al., 2009; Tichon et al., 2014). Although direct relationship with performance is only indirectly studied, results suggest that stressors, difficulty level, secondary cognitive tasks, emotional traits, and expertise can relate to performance accuracy and that higher stress related to poorer performance. The following themes also use non-invasive continuous measures of affect; however, the relationship between affect and performance becomes clearer.

Theme 2. The Effect of Negative-Activating Affect in Flight Performance

Theme 2 explores the role of negative-activating affective states, i.e. stress and anxiety, on flying performance. Stress and anxiety tend to be closely related (Harley & Azevedo, 2014). Both affective responses result in a negative subjective experience and increase in physiological arousal (Gross, 2015; Pekrun, 2019). However, anxiety is triggered by a specific stimulus and has shorter duration, whereas stress has a less specific trigger and has a longer duration (Harley et al., 2019a).

Stress and Flight Performance. An early study queried pilots (*n*=12) perception of stress across different flying scenarios, and predicted stress to impact their subsequent performance (Hart & Bortolussi, 1984). Pilots reported that challenging tasks, like taking off and landing, caused more stress, and stress increased with errors, having detrimental consequences on their subsequent performance. A more recent study similarly explored stress changes across specific segments in a flight simulator, assessing stress through HRV (Cao et al., 2019). Cao et al. 2019 found that pilots (*n*=30) had higher stress levels (lower HRV) than healthy adults. Pilots experienced more stress during challenging segments, like takeoff, compared to easy segments, like cruise control. Easy segments were granted a higher performance score according to instructors. Together, these two studies confirm that pilots experience more stress during challenging flight segments, which can lead to less accurate performance (Cao et al., 2019; Hart & Bortolussi, 1984).

Additional studies explored stress in relation to task demands. Vine et al. (2015) explored pilots' (*n*=16) appraisals of the task, attentional control, and performance in a flight simulator (Vine et al., 2015). A stress response was created by inducing an engine failure. Perceptions of task demand and personal resources positively correlated with performance accuracy, assessed by instructor scores and deviations from instructed flight. Pilots who reported that they perceived tasks as more challenging had lower attentional control (inferred from gaze), which associated to more stress. Interestingly, results show a direct link between cognitive appraisals and performance, explained by poorer attentional control; however, stress per se was not measured. Vallès-Català et al. (2021) investigated stress by examining pilots performing highly demanding tasks in a simulator. Instructors categorized demand level based on their perception of workload and assessed pilot performance as low, standard, and high. Changes in performance were assessed from task to task, visualizing improvement or decrease in performance. When student pilots' (*n*=41) performance improved, they experienced more stress, inferred from increases in EDA. When analyzing transitions from better to poorer performance, participants who a had higher overall score experienced less stress, compared to those with low scores. This study demonstrates that stress can be beneficial for performance when paired with better performance across task; however, if student pilots have continuous poor performance, experiencing more stress might be detrimental.

Cumulatively, these research findings demonstrate that pilots experience more stress than other healthy adults (Cao et al., 2019), highlighting the need to understand the impact of stress in pilots. The results demonstrate that more challenging flight segments cause an increase in stress, and less accurate performance, and the experience of stress is tightly attached to pilots' perception of the task (Cao et al., 2019; Hart & Bortolussi, 1984; Vallès-Català et al., 2021; Vine et al., 2015). Vallès-Català et al.'s (2021) findings demonstrate that stress levels should be evaluated along a threshold, similar to the Yerkes-Dodson law (1908): stress level within a normal distribution might require concentration to solve the task correctly, however, continuous poor performance might cause concern and thus high levels of stress through the task.

Anxiety During Flight Training. An early study by Smith and Melton (1978) compared participants' anxiety (*n*=16) using a ground trainer simulator compared to flying an airplane for an in-flight simulation. Participants who trained in the aircraft had higher anxiety (higher heart rate) compared to participants using the simulator, however, differences were not statistically significant, and performance did not differ.

More recent studies explored the impact of inducing anxiety on cognitive processing, and performance. The anxiety-induction technique was developed by Allsop and Gray (2014). This technique consisted of an ego-threatening depiction where participants were told that the best performance would obtain a monetary price, and the worse performance would be used as an example of bad performance. Allsop and Gray (2014) used this technique to understand the impact of anxiety on gaze patterns and performance in a computer-based simulation. The experimental group (who received anxiety-inducing depiction) had an increase in anxiety according to self-reports, heart rate, and visual scanning entropy. Performance in experimental and control groups was constant (*n*=20 participants). When analyzing the experimental group alone, authors found a correlation between distance from the ideal path and visual-scanning entropy, indicating that increased anxiety suggests a negative effect on attentional control.

A few studies reproduced this finding using the same anxiety induction technique. Gray et al. (2016) explored changes in attention finding that participants' (*n*=80) heart rate increased during the anxiety condition, and they were farther from the instructed path, showing poorer performance. Additionally, participants were more accurate shifting attention, but less accurate focusing and dividing attention in the anxiety condition. Another study using this technique, but instead of interpreting it as inducing anxiety, examined the intervention from a social stressor perspective due to the competitive nature of the task, instead of interpreting it as inducing anxiety (Hidalgo-Munoz et al., 2018). The intervention did not show differences in student pilots' (*n*=21) deviation from instructed flight path. In the non-anxious condition, student pilots were less accurate at detecting numbers according to simple or complex rules in the secondary cognitive task. Heart rate decreased across time, but HRV increased, implying less anxiety, explained due to task habituation.

The last study using Allsop and Gray's (2014) technique also interpreted it as a social stressor. This study aimed to create workload and social stress conditions to explore these concepts independently and understand their impact on performance (Causse et al., 2024). Pilots (*n*=20) flew two scenarios from takeoff to landing in a simulator. Pilots' workload was induced by adding a secondary task, consisting of identifying numbers, creating low and high workload conditions. Social stress was induced creating a sense of competition following Allsop and Gray's method, in the non-stress condition pilots would fly without additional instructions. Results showed that flying performance was sustained regardless of workload or stress conditions. However, pilots identified numbers faster during the stress condition.

Using a different approach, a study explored the relationship between emotional intelligence and performance mediated by state anxiety (Dai et al., 2019). Participants (*n*=90) responded to questionnaires to assess their emotional intelligence profile and state anxiety. Results did not show significant relations between anxiety, tension, and flying performance, as scored by instructors. Higher emotional intelligence with lower state-anxiety related to better performance.

The studies exploring anxiety demonstrate that this emotion can increase according to the physical and social environment: anxiety is higher when flying an airplane in an in-flight simulation, and when creating a competitive environment (Allsop & Gray, 2014; Causse et al., 2024; Gray et al., 2016; Hidalgo-Munoz et al., 2018; Smith & Melton, 1978). The impact of anxiety on performance is not conclusive. In some studies performance did not differ despite pilots' anxiety levels (Allsop & Gray, 2014; Causse et al., 2024; Dai et al., 2019; Hidalgo-Munoz et al., 2018; Smith & Melton, 1978). However, in two studies, experiencing more anxiety was detrimental for performance (Dai et al., 2019; Gray et al., 2016). In some cases anxiety is related to pilots' attention, indicating that anxiety improves performance in cognitive tasks that require attention shifts, but negatively affects performance where tasks require focusing (Allsop & Gray, 2014; Causse et al., 2024; Gray et al., 2016; Hidalgo-Munoz et al., 2018).

Theme 3. Social Dynamics Induce Affect in Simulated Flight

When flights are performed by more than one member, aircrew coordination is key for successful flight. Moreover, training is situated in a specific context, guided by social interactions (Sawyer & Greeno, 2012). Thus, theme 3 emphasizes how social dynamics could explain the relationship between affect and flying performance.

The first social pattern refers to the “teacher-student” dynamics. According to these studies, instructors’ feedback can induce affective responses in student pilots. For instance, the type of verbal feedback of instructors (n=6) can impact pilot students (n=12) stress when using a flight simulator (Krahenbuhl et al., 1981). Students who received praise had lower stress. Moreover, flight difficulty and instructor assistance can influence student pilots’ stress while practicing flying manoeuvrers (Skibniewski et al., 2015). Results demonstrated that the flight itself increased stress, as student pilots (n=59) had higher HRV during flight compared with pre and post stages. Stress was higher while performing high-difficulty manoeuvres (acrobatics) versus simple circle flight. Instructors’ presence did not affect stress levels. These results demonstrate that instructor’s presence might be beneficial when they provide positive and constructive feedback to trainees, by reducing stress. Unfortunately, neither of the studies presented results regarding performance accuracy.

Another group of studies explored peer-interactions to explain affect. A first study explored pilot students (n=28) interacting in the role of pilot and co-pilot in a regular flight (Wang et al., 2016). Laughter, as a behavioural expression of joy, was examined. Crew effectiveness, as scored by instructors, was related to task complexity and the time dyads had known each other. Pilot and co-pilot's independent laughter did not relate to performance; however, shared laugher negatively predicted performance. Shared laughter might imply a stronger social connection, but less task concentration, resulting in poorer performance. A second study explored temporal patterns of stress in student dyads (n=14) performing a complex landing in a simulator (Sassenus et al., 2022). Stress was analyzed from verbal responses. Results showed that stress was present 67% of the time. Stress triggers varied according to role: pilots got stressed by what the co-pilot said, whereas co-pilots got stressed by task demands.

The studies in Theme 3 re-emphasize that task difficulty is significantly related to pilots' performance and affective experience. Particularly, higher difficulty is related to poorer performance and higher stress for individual pilots and co-pilots (Sassenus et al., 2022; Skibniewski et al., 2015; Wang et al., 2016). Interestingly, in shared flying, students in the role of pilot were more stressed by what the co-pilot said (Sassenus et al., 2022), and performance was poorer when pilot and co-pilot laughed together, which might indicate higher social connection, but lower task concentration (Wang et al., 2016). Yet, this might be explained by having less experience, the interaction between licenced pilots might differ. Regarding student pilot-instructor interaction, instructors' presence seem to be beneficial only when providing positive and constructive feedback (Krahenbuhl et al., 1981; Skibniewski et al., 2015). Social dynamics impact pilots' affect, likely influencing their performance accuracy.

Theme 4. Unusual Flying Scenarios Impact Affect and Performance

Drastic changes in stress and emotions alter subsequent flying performance, making pilots prone to faulty decision making after unexpected events (Landman et al., 2017). Studies included in Theme 4 explore pilots' capacity to re-balance affective reactions and performance after experiencing unusual events on the aircraft.

Krahenbuhl et al. (1978) used simulators to test stress effects on student pilots (*n*=20) recovering from a spin manoeuvre. The intervention group had two opportunities to practice in the simulator, whereas the control group performed spin recoveries without previous practice. Findings proved that spin recovery simulation was stressful for both groups, observed in neuroendocrine changes analyzed in urine before and after the intervention. The control group had more emotional arousal, inferred from higher epinephrine levels, whereas the experimental group had higher stress, inferred from higher norepinephrine. Performance, assessed by the time following the instructed path was not different across groups. Likely, two opportunities to practice were not sufficient to create sufficient changes in affect arousal nor performance. Similarly, a more recent study tested practicing recovering from unusual aircraft changes (Koglbauer et al., 2011). In this case, pilots (*n*=29) started with an in-flight simulation in an airplane to create a sense of authenticity. Then, the experimental group practiced nine recovery scenarios in a simulator, and the control group practiced only two times. The experimental group significantly lowered their time on task, according to instructors' evaluation. Both groups reported an increase in positive emotions, with higher excitement, pride, and enthusiasm. Additionally, heart rate increased, and continuous levels of EDA decreased over time. The results of these two studies show that more than two practice sessions might be necessary to improve spin recovery performance; however, affective responses might be similar despite having more or less practice (Koglbauer et al., 2011; Krahenbuhl et al., 1978).

Despite technological advancement in flight automation, pilots still need to respond efficiently to unexpected events (Landman et al., 2017). Therefore, a growing interest in flight training is understanding the impact of startle and surprise in performance accuracy, as shown by the following studies. Kinney and O'Hare (2020) explored pilots' (*n*=22) startle reactions to unexpected events, measured by eye movements and heart rate. Results showed that pilots had stronger startle and surprise reactions during unexpected compared to expected events, observed in a significant increase in heart rate and pupil dilatation. Moreover, pilots landed safely 54.5% of unexpected events, and 100% of expected events. Descriptive findings show that participants who landed safely in unexpected events had a higher heart rate, smaller pupil dilatation, and spent less time looking at the cockpit display which is interpreted as having a more controlled arousal response and lower cognitive workload.

The results of these studies demonstrate that unexpected events significantly decrease performance accuracy (Kinney & O'Hare, 2020), and impact pilots' affect, increasing stress, anxiety, startle, and surprise immediately after the unexpected event (Kinney & O'Hare, 2020; Krahenbuhl et al., 1978). Practice might play a role in pilots' affect and performance. More than two rounds of practice are necessary to improve performance (Koglbauer et al., 2011; Krahenbuhl et al., 1978), however, only two opportunities to practice might provoke positive-activating emotions, which are beneficial for re-engaging in similar tasks in the future (Koglbauer et al., 2011).

Theme 5. Interventions to Manage Performance and Affect

The high-stakes characteristics of aviation makes pilots prone to experiencing stress (DeMaria et al., 2010). Excessive stress can negatively impact perception of a situation, producing incapacity to take appropriate actions to solve the situation (Landman et al., 2017). To overcome these difficulties, studies in theme 5 created management interventions to reduce detrimental consequences on performance.

Three studies designed simulator interventions to help pilots recover from spatial disorientation, defined as a mistaken sense of plane position in relation to the Earth's surface, threatening flying safety (Kallus & Tropper, 2004; Kang et al., 2020). A first intervention explored differences in stress responses when receiving training for recovering from unusual scenarios with a moving versus a static simulator (Kallus & Tropper, 2004). Results showed that pilots (*n*=42) who received training with movement were scored higher by instructors and self-reported a better performance. Stress was measured with self-reports, but results were not reported. In a follow-up study (Tropper et al., 2009) pilots (*n*=25) were divided into three intervention groups: awareness, awareness and orientation training, and control group, who only performed free flight. Awareness intervention referred to exposing the pilots to unusual scenarios in the simulator. The second group received instructors' guidance of how to recover from the scenario (orientation), in addition to the awareness intervention. The control group was scored significantly lower by instructors, had more crashes, and higher stress, inferred from heart rate. Mixed orientation and awareness group obtained better scores overall, and they experienced more HRV, interpreted as being more skilled to relax and return to a basal state between tasks.

Another study used verbal reports to reduce stress and improve performance in spatial disorientation tasks (Kang et al., 2020). Pilots (*n*=30) who verbalized the procedure they were going to take were scored significantly higher by instructors, had higher HRV, and reported less stress, compared to those who did not verbalize. Generally, the results of these studies show that using moving vs. static simulators, awareness techniques, instructors guidance, and verbalizing actions improve performance in spatial disorientation tasks and such interventions reduce stress (Kallus & Tropper, 2004; Kang et al., 2020; Tropper et al., 2009).

A second set of studies implemented affect management interventions. A study tested a mnemonic technique for managing startle in unexpected events (Landman et al., 2020). The mnemonic was COOL: calm down, observe, outline, lead. Pilots (*n*=24) were divided into intervention and control group. Results show that 60% of the pilots in the control group observed, despite not being trained to use the mnemonic technique, whereas 89.6% of the intervention group implemented all the steps in the mnemonic. Effects on performance were variable. Right after a surprising event, the experimental group scored significantly lower in their immediate response to ensure safe flight; both groups had equal ability to assess the cause, but once pilots in the experimental group recovered from a surprise, they obtained better scores on actions taken to continue a safe flight.

Another study explored a stress mitigation technique for concentrating when doing an in-flight simulation, despite having external stressors (cold pressor on the foot) (McClernon et al., 2011). The technique consisted of reading instructions to encourage pilots to concentrate on the task and remain relaxed. Participants (n=30) were divided in intervention and control group, who did not receive instructions. The intervention group had less movement in their flight, being more precise, compared to the control group. Both groups reported a decrease in subjective experience of stress.

These studies show that techniques for monitoring performance and affect can improve flying accuracy, to guarantee safe flights despite abnormalities (Landman et al., 2020). Specialized training might successfully diminish detrimental consequences of affective reactions linked to unusual events (Kallus & Tropper, 2004; Kang et al., 2020; Tropper et al., 2009). Through intentional practice in authentic environments, instructors' guidance, verbalizing actions, and instruction on emotion regulation may help pilot trainees to experience emotions in a functional manner, and recover from them, resulting in better performance (Kallus & Tropper, 2004; Kang et al., 2020; Tropper et al., 2009). Findings suggest that pilots who received tools for self-regulating affect can have better control over their flight (Landman et al., 2020; McClernon et al., 2011).

Conclusions

Our results indicate that since the beginning of the 21st century there is a growing interest in understanding the relationship between affect and performance when training with flight simulations. The technological advances in aviation make simulations desirable for pilot training (Cross et al., 2023; Lajoie & Poitras, 2023; Salas et al., 2010). Our results show that simulations can induce affect in an authentic manner, allowing pilot trainees to familiarize themselves with their reactions, and regulate themselves, to achieve successful performance (Gordon et al., 2010; Lajoie, 2017).

The rationale for examining pilot affect is that intense emotional responses can distract pilots and loosing concentration could result in errors (Dismukes, 2010; Mathavara et al., 2022). Most of the studies explored affect to understand how to improve performance accuracy. Theme 1 demonstrated that multimodal measures (behavioural, physiological, and self-reported) can be used to explore the temporal nature of affect in a continuous and non-invasive manner, such that pilots performance, as a concurrent processes, is not interrupted (Drinkwater et al., 1968; Gaetan et al., 2015; Li & Lajoie, 2021; Rosa et al., 2021, 2022; Silva et al., 2009; Tichon et al., 2014). Theme 2 revealed that negative-activating affective responses, like stress and anxiety, tend to be detrimental for flying accuracy (Cao et al., 2019; Dai et al., 2019; Gray et al., 2016; Hart & Bortolussi, 1984; Vallès-Català et al., 2021; Vine et al., 2015). However, the relationship between negative-activating responses and performance should be interpreted according to a threshold of intensity: negative-activating affect can have a beneficial or standard impact of performance, whereas high negative-activating affect can be detrimental (Vallès-Català et al., 2021; Yerkes &

Dodson, 1908). Namely, anxiety is beneficial for tasks that require shifting attention, but detrimental when the tasks demands focus (Allsop & Gray, 2014; Causse et al., 2024; Gray et al., 2016; Hidalgo-Munoz et al., 2018). Although we divided the articles following a theoretical distinction, a limitation is that stress and anxiety are closely related and, in many cases, overlap (Ahn et al., 2023; Fink, 2006). More research could explore theoretical and experimental specifications to identify the differences between stress and anxiety (Cao et al., 2019)

Theme 3 showed that social dynamics influence pilots' affect. Co-pilots can be more stressed by task demands, while pilots get stressed by the co-pilots' messages (Sassenus et al., 2022); furthermore a high social connection can decrease concentration and performance (Wang et al., 2016). Instructors presence was beneficial for performance only when they provide positive feedback (Krahenbuhl et al., 1981; Skibniewski et al., 2015).

Theme 4 showed that unusual events increase stress, anxiety, startle, and surprise, and tend to be detrimental for performance accuracy (Kinney & O'Hare, 2020; Koglbauer et al., 2011; Krahenbuhl et al., 1978). Interestingly, opportunities to practice might increase positive-activating emotions, which is beneficial for learners motivation (Koglbauer et al., 2011; Pekrun, 2019).

The studies in the last theme demonstrated that pilots' affective responses and performance can improve with interventions. Namely, performance can improve through intentional practice in authentic environments, instructors' guidance, and verbalizing actions (Kallus & Tropper, 2004; Kang et al., 2020; Tropper et al., 2009). Simple emotion regulation techniques, like remembering a mnemonic to manage surprise reactions after unexpected events, can help pilots to regulate their affective responses, and consequently improve performance (Landman et al., 2020; McClernon et al., 2011).

This systematic review had some limitations. The selected databases concentrate on educational and psychological research, restraining our access to articles in other domains. Although it was not an exclusion criterion, studies included were all written in English, limiting access to a wider range of perspectives reported in less dominant languages. Lastly, we did not consider the theoretical quality as a criterion for including studies, which resulted in having inconclusive or contradicting findings. Future studies should carefully align theoretical assumptions with methodological approaches to better understand the role of affect in flying performance. In the future, we will conduct a second literature review discussing the advantages and disadvantages of using different measures of affect and flying performance.

This systematic review demonstrates the relationships between affect and performance in flight training. A main finding is that unpredictable events can result in intense affective reactions that can distract pilots, resulting in errors. Our insights suggest some future directions. We argue that all affective responses have a function, but current research mainly focuses on negative and activating emotions (stress, anxiety, and surprise). Future research can explore more affective responses like positive emotions, i.e., joy and pride, that can improve performance (Pekrun, 2019). More research will be needed to explore the relationship between flight automation and cognitive effort for pilots. If cognitive workload is decreased there may be a lack of stimulation that can cause pilots to experience more deactivating emotions, having detrimental consequences on performance (Pekrun, 2019). Negative-deactivating emotions like boredom and hopelessness are confirmed to lead to task disengagement, causing poorer performance (D'Mello & Graesser, 2012); but the functionality of positive-deactivating emotions, like relaxation and relief, remains to be explored.

References

Ahn, B., Maurice-Ventouris, M., Bilgic, E., Yang, A., Lau, C. H.-H., Peters, H., Li, K., Chang-Ou, D., & Harley, J. M. (2023). A scoping review of emotions and related constructs in simulation-based education research articles. *Advances in Simulation*, *8*(1), 22. https://doi.org/10.1186/s41077-023-00258-z

Allsop, J., & Gray, R. (2014). Flying under pressure: Effects of anxiety on attention and gaze behavior in aviation. *Journal of Applied Research in Memory and Cognition*, *3*(2), 63–71. https://doi.org/10.1016/j.jarmac.2014.04.010

American Psychological Association. (n.d.). Affect. In *APA dictionary of psychology*. Retrieved December 5, 2023, from https://dictionary.apa.org/affect

Bent, J., & Chan, K. (2010). Flight training and simulation as safety generators. In E. Salas & D. Maurino (Eds.), *Human Factors in Aviation* (2nd ed., pp. 293–334). Academic Press. https://doi.org/10.1016/B978-0-12-374518-7.00010-9

Braithwaite, J. J., Watson, D. G., Jones, R., & Rowe, M. (2015). *A guide for analysing electrodermal activity (EDA) and skin conductance responses (SCRs) for psychological experiments*. SAAL, Behavioural Brain Sciences Centre, University of Birminghan. https://www.birmingham.ac.uk/Documents/college-les/psych/saal/guide-electrodermal-activity.pdf

Canadian Aviation Regulations. (2022, December 21). *Commercial air services*. Justice laws website - Government of Canada. https://laws-lois.justice.gc.ca/eng/regulations/sor-96-433/page-83.html

Cao, X., MacNaughton, P., Cadet, L. R., Cedeno-Laurent, J. G., Flanigan, S., Vallarino, J., Donnelly-McLay, D., Christiani, D. C., Spengler, J. D., & Allen, J. G. (2019). Heart rate variability and performance of commercial airline pilots during flight simulations. *International Journal of Environmental Research and Public Health*, *16*(2), 237. https://doi.org/10.3390/ijerph16020237

Causse, M., Mouratille, D., Rouillard, Y., El Yagoubi, R., Matton, N., & Hidalgo-Muñoz, A. (2024). How a pilot's brain copes with stress and mental load? Insights from the executive control network. *Behavioural Brain Research*, *456*, 114698. https://doi.org/10.1016/j.bbr.2023.114698

Code of Federal Regulations. (2023, April 19). *14 CFR 91.1059 -- Flight time limitations and rest requirements: One or two pilot crews.* Electronic code of federal regulations. https://www.ecfr.gov/current/title-14/chapter-I/subchapter-F/part-91/subpart-K/subject-group-ECFRc17623c0e0be17e/section-91.1059

Cross, J. I., Boag-Hodgson, C. C., & Mavin, T. J. (2023). Measuring presence and situational awareness in a virtual reality flight simulator. *Aviation Psychology and Applied Human Factors*. https://doi.org/10.1027/2192-0923/a000250

Dai, J., Wang, H., Yang, L., & Wen, Z. (2019). Emotional intelligence and emotional state effects on simulated flight performance. *Aerospace Medicine and Human Performance*, *90*(2), 101–108. https://doi.org/10.3357/AMHP.5184.2019

Damasio, A. R. (2005). *Descartes' error: Emotion, reason, and the human brain*. Penguin Books.

DeMaria, S., Bryson, E. O., Mooney, T. J., Silverstein, J. H., Reich, D. L., Bodian, C., & Levine, A. I. (2010). Adding emotional stressors to training in simulated cardiopulmonary arrest enhances participant performance. *Medical Education*, *44*(10), 1006–1015. https://doi.org/10.1111/j.1365-2923.2010.03775.x

Dismukes, R. K. (2010). Understanding and analyzing human error in real-world operations. In E. Salas & D. Maurino (Eds.), *Human factors in aviation* (2nd ed., pp. 335–374). Academic Press. https://doi.org/10.1016/B978-0-12-374518-7.00011-0

D'Mello, S., & Graesser, A. (2012). Dynamics of affective states during complex learning. *Learning and Instruction*, *22*(2), 145–157. https://doi.org/10.1016/j.learninstruc.2011.10.001

Drinkwater, B. L., Flint, M. M., Clel, & Troy S. (1968). Somatic responses and performance levels during anticipatory physical-threat stress. *Perceptual and Motor Skills*, *27*(2), 539–552. https://doi.org/10.2466/pms.1968.27.2.539

Ekman, P. (2020). *What are emotions?* Paul Ekman Group. https://www.paulekman.com/universal-emotions/

European Commission. (2013, October 4). *Pilot and crew fatigue—Frequently asked questions*. Directorate-General for Communication, European Commission. https://ec.europa.eu/commission/presscorner/detail/en/MEMO_13_854

Eysenck, M. W., Derakshan, N., Santos, R., & Calvo, M. G. (2007). Anxiety and cognitive performance: Attentional control theory. *Emotion*, *7*(2), 336–353. https://doi.org/10.1037/1528-3542.7.2.336

Fink, G. (2016). Stress, definitions, mechanisms, and effects outlined: Lessons from anxiety. In G. Fink (Ed.), *Stress: Concepts, cognition, emotion, and behavior* (pp. 3–11). Academic Press. https://doi.org/10.1016/B978-0-12-800951-2.00001-7

Gaba, D. M. (2004). The future vision of simulation in health care. *Quality & Safety in Health Care*, *13 Suppl 1*(Suppl 1), i2-10. https://doi.org/10.1136/qhc.13.suppl_1.i2

Gaetan, S., Dousset, E., Marqueste, T., Bringoux, L., Bourdin, C., Vercher, J.-L., & Besson, P. (2015). Cognitive workload and psychophysiological parameters during multitask activity in helicopter pilots. *Aerospace Medicine and Human Performance*, *86*(12), 1052–1057. https://doi.org/10.3357/AMHP.4228.2015

Goetz, T., & Hall, N. C. (2014). Academic boredom. In R. Pekrun & L. Linnenbrink-Garcia (Eds.), *International handbook of emotions in education* (pp. 311–330). Routledge.

Gordon, J. A., Hayden, E. M., Ahmed, R. A., Pawlowski, J. B., Khoury, K. N., & Oriol, N. E. (2010). Early bedside care during preclinical medical education: Can technology-enhanced patient simulation advance the Flexnerian ideal? *Academic Medicine*, *85*(2), 370–377. https://doi.org/10.1097/ACM.0b013e3181c88d74

Gray, R., Gaska, J., & Winterbottom, M. (2016). Relationship between sustained, orientated, divided, and selective attention and simulated aviation performance: Training & pressure effects. *Journal of Applied Research in Memory and Cognition*, *5*(1), 34–42. https://doi.org/10.1016/j.jarmac.2015.11.005

Gross, J. J. (2015). Emotion regulation: Current status and future prospects. *Psychological Inquiry*, *26*(1), 1–26. https://doi.org/10.1080/1047840X.2014.940781

Harley, J. M. (2016). Measuring emotions: A survey of cutting edge methodologies used in computer-based learning environment research. In S. Y. Tettegah & M. Gartmeier (Eds.), *Emotions, technology, design, and learning* (pp. 89–114). Academic Press. https://doi.org/10.1016/B978-0-12-801856-9.00005-0

Harley, J. M., & Azevedo, R. (2014). Toward a feature-driven understanding of students' emotions during interactions with agent-based learning environments: A selective review.

International Journal of Gaming and Computer-Mediated Simulations, *6*(3), 17–34. https://doi.org/10.4018/ijgcms.2014070102

Harley, J. M., Pekrun, R., Taxer, J. L., & Gross, J. J. (2019a). Emotion regulation in achievement situations: An integrated model. *Educational Psychologist*, *54*(2), 106–126. https://doi.org/10.1080/00461520.2019.1587297

Hart, S. G. (2006). Nasa-Task Load Index (NASA-TLX): 20 years later. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *50*(9), 904–908. https://doi.org/10.1177/154193120605000909

Hart, S. G., & Bortolussi, M. R. (1984). Pilot errors as a source of workload. *Special Issue: Aviation Psychology*, *26*(5), 545–556.

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In P. A. Hancock & N. Meshkati (Eds.), *Human mental workload* (Vol. 52, pp. 139–183). Elsevier Science. https://doi.org/10.1016/S0166-4115(08)62386-9

Helmreich, R. L. (2010). Foreword. In E. Salas & D. Maurino (Eds.), *Human factors in aviation* (2nd ed., pp. xi–xii). Academic Press. https://doi.org/10.1016/B978-0-12-374518-7.00024-9

Hidalgo-Munoz, A. R., Mouratille, D., Matton, N., Causse, M., Rouillard, Y., & El-Yagoubi, R. (2018). Cardiovascular correlates of emotional state, cognitive workload and time-on-task effect during a realistic flight simulation. *International Journal of Psychophysiology*, *128*, 62–69. https://doi.org/10.1016/j.ijpsycho.2018.04.002

Hodhod, R., Fleenor, H., & Nabi, S. (2014). Adaptive augmented reality serious game to foster problem solving skills. *Workshop Proceedings of the 10th International Conference on Intelligent Environments*, 273–284. https://doi.org/10.3233/978-1-61499-411-4-273

Jarrell, A., Harley, J. M., Lajoie, S., & Naismith, L. (2017). Success, failure and emotions: Examining the relationship between performance feedback and emotions in diagnostic reasoning. *Educational Technology Research and Development*, *65*(5), 1263–1284. https://doi.org/10.1007/s11423-017-9521-6

Jarrell, A., & Lajoie, S. P. (2017). The regulation of achievements emotions: Implications for research and practice. Canadian Psychology/Psychologie Canadienne, 58(3), 276–287. https://doi.org/10.1037/cap0000119

Kallus, K. W., & Tropper, K. (2004). Evaluation of a spatial disorientation simulator training for jet pilots. *International Journal of Applied Aviation Studies*, *4*(1), 45–55. https://doi.org/10.1080/10508414.2011.556458

Kang, Y., Yun, M. H., & Kim, S. (2020). Verbal reports' influence on pilot flight performance and mental stress under spatial disorientation. *Aerospace Medicine and Human Performance*, *91*(12), 948–955. https://doi.org/10.3357/AMHP.5620.2020

Kinney, L., & O'Hare, D. (2020). Responding to an unexpected in-flight event: Physiological arousal, information processing, and performance. *Human Factors*, *62*(5), 737–750. https://doi.org/10.1177/0018720819854830

Koglbauer, I., Kallus, K., Braunstingl, R., & Boucsein, W. (2011). Recovery training in simulator improves performance and psychophysiological state of pilots during simulated and real visual flight rules flight. *The International Journal of Aviation Psychology*, *21*(4), 307–324. https://doi.org/10.1080/10508414.2011.606741

Krahenbuhl, G. S., Darst, P. W., Marett, J. R., Reuther, L. C., Constable, S. H., Swinford, M. E., & Reid, G. B. (1981). Instructor pilot teaching behavior and student pilot stress in flight training. *Aviation, Space, and Environmental Medicine*, *52*(10), 594–597.

Krahenbuhl, G. S., Marett, J. R., & Reid, G. B. (1978). Task-specific simulator pretraining and in-flight stress of student pilots. *Aviation, Space, and Environmental Medicine*, *49*, 1107–1110.

Lajoie, S. P. (2017). Learning science applications for research in medicine. In L. Lin & J. M. Spector (Eds.), *The sciences of learning and instructional design* (pp. 108–118). Routledge. https://doi.org/10.4324/9781315684444-8

Lajoie, S. P., & Azevedo, R. (2006). Teaching and learning in technology-rich environments. In P. A. Alexander & P. H. Winne (Eds.), *Handbook of educational psychology* (pp. 803–821). Lawrence Erlbaum Associates Publishers.

Lajoie, S. P., Pekrun, R., Azevedo, R., & Leighton, J. P. (2020). Understanding and measuring emotions in technology-rich learning environments. *Learning and Instruction*, *70*, 101272. https://doi.org/10.1016/j.learninstruc.2019.101272

Lajoie, S. P., & Poitras, E. (2023). Technology-rich learning environments: Theories and methodologies for understanding solo and group learning. In P. A. Schutz & K. R. Muis (Eds.), *Handbook of educational psychology* (4th ed., pp. 630–653). Routledge. https://doi.org/10.4324/9780429433726

Landman, A., Groen, E. L., van Paassen, M. M., Bronkhorst, A. W., Mulder, M., & Mulder, M. (2017). Dealing with unexpected events on the flight deck: A conceptual model of startle and surprise. *Human Factors*, *59*(8), 1161–1172. https://doi.org/10.1177/0018720817723428

Landman, A., van Middelaar, S. H., Groen, E. L., van Paassen, M. M., Bronkhorst, A. W., & Mulder, M. (2020). The effectiveness of a mnemonic-type startle and surprise management procedure for pilots. *The International Journal of Aerospace Psychology*, *30*(3–4), 104–118. https://doi.org/10.1080/24721840.2020.1763798

Li, T. (2020). *Multimodal affect assessment in aviation training* [Masters thesis, McGill University]. https://escholarship.mcgill.ca/concern/theses/6d5702130

Li, T., & Lajoie, S. (2021). Predicting aviation training performance with multimodal affective inferences. *International Journal of Training and Development*, *25*(3), 301–315. https://doi.org/10.1111/ijtd.12232

Makransky, G., Andreasen, N. K., Baceviciute, S., & Mayer, R. E. (2020). Immersive virtual reality increases liking but not learning with a science simulation and generative learning strategies promote learning in immersive virtual reality. Journal of Educational Psychology, 113, 719–735. https://doi.org/10.1037/edu0000473

Martin, W. L., Murray, P. S., Bates, P. R., & Lee, P. S. Y. (2015). Fear-potentiated startle: A review from an aviation perspective. *The International Journal of Aviation Psychology*, *25*(2), 97–107. https://doi.org/10.1080/10508414.2015.1128293

Martín-Gutiérrez, J., Mora, C. E., Añorbe-Díaz, B., & González-Marrero, A. (2017). Virtual Technologies Trends in Education. *Eurasia Journal of Mathematics, Science and Technology Education*, *13*(2), 469–486. https://doi.org/10.12973/eurasia.2017.00626a

Mathavara, K., & Ramachandran, G. (2022). Role of human factors in preventing aviation accidents: An insight. In Z. A. Ali & D. Cvetković (Eds.), *Aeronautics—New Advances*. IntechOpen. https://doi.org/10.5772/intechopen.106899

McClernon, C. K., McCauley, M. E., O'Connor, P. E., & Warm, J. S. (2011). Stress training improves performance during a stressful flight. *Human Factors*, *53*(3), 207–218. https://doi.org/10.1177/0018720811405317

McClernon, C. K., & Miller, J. C. (2011). Variance as a measure of performance in an aviation context. *The International Journal of Aviation Psychology*, *21*(4), 397–412. https://doi.org/10.1080/10508414.2011.606765

Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. G. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *BMJ*, *339*, b2535. https://doi.org/10.1136/bmj.b2535

Morris, C. S., Hancock, P. A., & Shirkey, E. C. (2004). Motivational effects of adding context relevant stress in PC-based game training. *Military Psychology*, *16*(2), 135–147. https://doi.org/10.1207/S15327876MP1602_4

Muñoz, J. E., Montoya, M. F., & Boger, J. (2022). From exergames to immersive virtual reality systems: Serious games for supporting older adults. In M.-A. Choukou & S. Syed-Abdul (Eds.), *Smart home technologies and services for geriatric rehabilitation* (pp. 141–204). Academic Press. https://doi.org/10.1016/B978-0-323-85173-2.00011-4

Ouzzani, M., Hammady, H., Fedorowicz, Z., & Elmagarmid, A. (2016). Rayyan—A web and mobile app for systematic reviews. *Systematic Reviews*, *5*(1), 210. https://doi.org/10.1186/s13643-016-0384-4

Page, M. J., Moher, D., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., … McKenzie, J. E. (2021). PRISMA 2020 explanation and elaboration: Updated guidance and exemplars for reporting systematic reviews. *BMJ*, *372*, n160. https://doi.org/10.1136/bmj.n160

Parong, J., & Mayer, R. E. (2020). Cognitive and affective processes for learning science in immersive virtual reality. *Journal of Computer Assisted Learning*, *1*(16). https://doi.org/10.1111/jcal.12482

Pekrun, R. (2019). Achievement emotions: A control-value theory perspective. In R. Patulny, A. Bellocchi, R. Olson, S. Khorana, J. McKenzie, & M. Peterie (Eds.), *Emotions in late modernity* (pp. 142-157). Routledge.

Pekrun, R., & Perry, R. P. (2014). Control-value theory of achievement emotions. In R. Pekrun & L. Linnenbrink-Garcia (Eds.), *International handbook of emotions in education* (pp. 120–141). Routledge.

Pilote, B., & Chiniara, G. (2019). The many faces of simulation. In G. Chiniara (Ed.), *Clinical simulation* (2nd ed., pp. 17–32). Academic Press. https://doi.org/10.1016/B978-0-12-815657-5.00002-4

Roh, Y. S., & Jang, K. I. (2017). Survey of factors influencing learner engagement with simulation debriefing among nursing students. *Nursing & Health Sciences*, *19*(4), 485–491. https://doi.org/10.1111/nhs.12371

Rosa, E., Gronkvist, M., Kolegard, R., Dahlstrom, N., Knez, I., Ljung, R., & Willander, J. (2021). Fatigue, emotion, and cognitive performance in simulated long-duration, single-piloted flight missions. *Aerospace Medicine and Human Performance*, *92*(9), 710–719. https://doi.org/10.3357/AMHP.5798.2021

Rosa, E., Lyskov, E., Gronkvist, M., Kolegard, R., Dahlstrom, N., Knez, I., Ljung, R., & Willander, J. (2022). Cognitive performance, fatigue, emotional, and physiological strains in simulated long-duration flight missions. *Military Psychology*, *34*(2), 224–236. https://doi.org/10.1080/08995605.2021.1989236

Rudland, J. R., Golding, C., & Wilkinson, T. J. (2020). The stress paradox: How stress can be good for learning. *Medical Education*, *54*(1), 40–45. https://doi.org/10.1111/medu.13830

Russell, J. A. (2003). Core affect and the psychological construction of emotion. *Psychological Review*, *110*(1), 145–172. https://doi.org/10.1037/0033-295X.110.1.145

Salas, E., Maurino, D., & Curtis, M. (2010). Human factors in aviation: An overview. In E. Salas & D. Maurino (Eds.), *Human factors in aviation* (2nd ed., pp. 3–19). Academic Press. https://doi.org/10.1016/B978-0-12-374518-7.00001-8

Sassenus, S., Van den Bossche, P., & Poels, K. (2022). When stress becomes shared: Exploring the emergence of team stress. *Cognition, Technology & Work*, *24*, 537–556. https://doi.org/10.1007/s10111-022-00698-z

Sawyer, R. K., & Greeno, J. G. (2012). Situativity and learning. In P. Robbins & M. Aydede (Eds.), *The Cambridge handbook of situated cognition* (pp. 347–367). https://doi.org/10.1017/CBO9780511816826.018

Silva, D. C., Vinhas, V., Reis, L. P., & Oliveira, E. (2009). Biometric emotion assessment and feedback in an immersive digital environment. *International Journal of Social Robotics*, *1*(4), 307–317. https://doi.org/10.1007/s12369-009-0029-z

Skibniewski, F. W., Dziuda, L., Baran, P. M., Krej, M. K., Guzowski, S., Piotrowski, M. A., & Truszczynski, O. E. (2015). Preliminary results of the LF/HF ratio as an indicator for estimating difficulty level of flight tasks. *Aerospace Medicine and Human Performance*, *86*(6), 518–523. https://doi.org/10.3357/AMHP.4087.2015

Smith, R. C., & Melton, C. E. (1978). Effects of ground trainer use on the anxiety of students in private pilot training. *Aviation, Space, and Environmental Medicine*, *49*, 406–408.

Tichon, J. G., Mavin, T., Wallis, G., Visser, T. A. W., & 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.

Tropper, K., Kallus, K. W., & Boucsein, W. (2009). Psychophysiological evaluation of an antidisorientation training for visual flight rules pilots in a moving base simulator. *The International Journal of Aviation Psychology*, *19*(3), 270–286. https://doi.org/10.1080/10508410902983912

Vallès-Català, T., Pedret, A., Ribes, D., Medina, D., & Traveria, M. (2021). Effects of stress on performance during highly demanding tasks in student pilots. *The International Journal of Aerospace Psychology*, *31*(1), 43–55. https://doi.org/10.1080/24721840.2020.1841564

Vine, S. J., Uiga, L., Lavric, A., Moore, L. J., Tsaneva-Atanasova, K., & Wilson, M. R. (2015). Individual reactions to stress predict performance during a critical aviation incident. *Anxiety, Stress & Coping: An International Journal*, *28*(4), 467–477.

Wang, L., Doucet, L., Waller, M., Sanders, K., & Phillips, S. (2016). A laughing matter: Patterns of laughter and the effectiveness of working dyads. *Organization Science*, *27*(5), 1142–1160. https://doi.org/10.1287/orsc.2016.1082

Weingarten, N. C. (2005). History of in-flight simulation at general dynamics. *Journal of Aircraft*, *42*(2), 290–298. https://doi.org/10.2514/1.4663

Yerkes, R. M., & Dodson, J. D. (1908). The relation of strength of stimulus to rapidity of habit formation. *Journal of Comparative Neurology & Psychology*, *18*, 459–482. https://doi.org/10.1002/cne.920180503

Bridging Text

Chapter 2 presents a systematic literature review synthesizing previous research that has explored the relationship between pilots' affect and flying performance when using simulated flying tasks, aiming to respond to the research question: What is the relation between affect and flying performance when training with flight simulations? The empirical articles were grouped in themes according to the manner they explain the relationship between affect and performance, explaining this relationship based on: 1) accounting for affect and performance as continuous and concurrent processes, 2) the effect of negative-activating affect (i.e., stress and anxiety) in flight performance, 3) how social dynamics induce affect in simulated flight, 4) unusual flying scenarios impacting affect and performance and 5) using interventions to manage performance and affect. This systematic review revealed that different expressions of affect are correlated to distinctive performance outcomes when using flight simulations.

The findings of the literature review showed that flight training research focuses mainly on specific expressions of affect such as anxiety, stress, and surprise according to specific environmental characteristics, such as difficulty level, social interactions, and type of interventions. Chapter 3 attempts to contribute to the literature by empirically exploring a wider range of discrete emotions, and the dynamic patterns of emotions that beginner pilot trainees experience during simulated flight. Particularly, Chapter 3 explores differences in pilot trainees flying performance and emotional expressions across three subsequent training phases and three difficulty levels (low, medium, and high).

Chapter 3. Manuscript 2

Flight Emotions Unleashed: Navigating Training Phases and Difficulty Levels in Simulated Flying

Ruiz-Segura, A. Law, A., Jennings, S., Bourgon, A., Churchill, E., & Lajoie, S. (under review). Flight emotions unleashed: Navigating training phases and difficulty levels in simulated flying. *The Journal of Computer Assisted Learning.*

Abstract

Background: Flying accuracy is influenced by pilots' affective reactions to task demands. A better understanding of task-related emotions and flying performance is needed to enhance pilot training. **Objective:** Understand pilot trainees' performance and emotional dynamics (intensity, frequency, and variability) based on training phase and difficulty level in a flight simulator. **Methods:** Twenty-three volunteers performed basic flight maneuvers. Trials were divided into three phases: Introduction (trials 1-7), Session A (trials 8-15), and Session B (trials 16-22). Three task difficulty levels were implemented (low, medium, and high). Flying performance was evaluated using root-mean-square-error and expert ratings. Emotional intensity was inferred from physiological (electrodermal activity) and behavioral (facial expressions) emotional responses. Emotional variability was calculated to understand fluctuations among multiple emotions. Emotional responses were mapped into task-relevant emotions, like sadness with boredom, and fear with anxiety. **Results and conclusions:** The most frequent facial expressions neutral and anger were interpreted as deep focus states. Flying performance and emotional dynamics varied across training phases and difficulty levels. During Introduction, performance was less accurate, and emotions were less frequent. During Session A, performance improved while participants experienced more physiological arousal and emotional variability. During session B, performance was the most accurate, and participants expressed more boredom. In high-difficulty tasks, performance was the least accurate, participants expressed emotions with more frequency, more variability, and higher physiological arousal. Future studies can use simulated flying tasks for trainees to familiarize with their emotional reactions to task demands expecting to improve training outcomes.

Keywords: Simulation, Flight, Training, Emotion, Facial expression, Electrodermal activity

Introduction

Aviation is a safety-critical and high-stakes profession in which errors can have significant detrimental consequences on flight safety (Baumann et al., 2011). The aviation industry invests great efforts into diminishing human error through the improved design of flight deck systems as well as effective training of pilots (Code of Federal Regulations, 2023). Simulations in aviation have been a fundamental training technique since World War II (Hamman, 2004; Herrera-Aliaga & Estrada, 2022). Simulations are defined as “a technique, not a technology, to replace or amplify real experiences with guided experiences that evoke or replicate substantial aspects of the real world in a fully interactive manner” (Gaba, 2004, pp. i2). In this text, the term “simulator” will be used to refer to the devices that physically mimic airplanes (Gaba, 2004; Lajoie, 2021).

From an educational perspective, simulations allow learners to deliberately practice skills in a safe and authentic environment, allowing them to be prepared to transfer those skills to the real world (Azher et al., 2023; Ericsson, 2006; Lajoie, 2021). Previous research has confirmed that simulations evoke authentic emotions, like those experienced using real airplanes (Koglbauer et al., 2011; Skibniewski et al., 2015). Using simulations might be especially beneficial for pilot trainees to familiarize themselves with the stakes of flying an airplane, and habituate to their emotional reactions in a safe context and controlled environment (Vine et al., 2015). However, the traditional measurements of emotions are conflicting with the objectives of flight training. Emotions are traditionally measured using questionnaires (Harley et al., 2016; Duffy et al., 2016). But a common critique of using questionnaires is that they can be distracting, and the reporting of the emotion is differed from the moment that the emotion was experienced (Harley et al., 2016). Moreover, the aviation industry is recognized for investing efforts into preventing errors that may be due to distractions and faulty decision making and intense emotions experienced during real flights (Hamman, 2004). Therefore, the objective of this study is to explore performance and emotional changes while performing maneuvers in a flight simulator across training phases and difficulty levels.

Previous research in aviation has been interested in understanding the relationships between affective responses (i.e., mood, stress, emotions) on flying performance when using flight simulators. The most common theoretical frameworks used to understand affect in flight training include stress and coping theory (Lazarus & Folkman, 1984) (e.g., Sassenus et al., 2022; Tichon et al., 2014), situational awareness theory (Endsley, 2000) (e.g., Gray et al., 2016; Kinney & O'Hare, 2020) and the circumplex model of affect (Russell, 1980, 2003) (e.g., Rosa et al., 2021, 2022), the latter being the only of these frameworks that focuses on emotions. In this manuscript, moods are excluded since these have a low physiological arousal and longer lasting affective states, persisting for days, and thus the relation to performance in a standard commercial flight might be unclear (lasting a maximum of 14 to 16 hours) (Code of Federal Regulations, 2023; European Commission, 2013; Gross, 2015). Stress is excluded since it occurs in taxing situations, being experienced as displeasing, and it is not stimuli-specific; in this manuscript, we attempt to understand a range of positive and negative emotions (DeMaria et al., 2010; Landman et al., 2017). Since emotions are context and stimuli specific (Gross, 2015), this study proposes to use a guiding theory that focuses on emotions that occur during the training process. We follow Pekrun's control-value theory of achievement emotions (2019), which has been used to understand training of pilots (T. Li & Lajoie, 2021) and other high-stakes professions like nursing (Azher et al., 2023; Harley et al., 2023) and medicine (Artino et al., 2012; Lajoie et al., 2023; Nomura et al., 2021).

Emotions and Achievement Performance

Emotions are defined as quick-changing affective states that last from seconds to minutes and have a clear object focus (Harley et al., 2019a). Emotions are directly connected to a specific stimulus, rather than being a broad feeling as is the case of mood (i.e., feeling down) or stress (Gross, 2015). Moods differ from emotions since they are longer lasting, sustaining for days, and the feeling is overall less arousing (Gross, 2015). Stress is characterized as a displeasing feeling experienced during taxing situations, whereas emotions have a range of pleasantness (Gross, 2015). Therefore, in this manuscript it is argued that emotions might provide more rich information about the range of pleasing and displeasing feelings that are triggered during a flying task.

Achievement emotions can be classified according to their valence and arousal (Pekrun, 2019). Valence represents the subjective feeling of pleasantness (positive or negative), and arousal refers to the physiological activation or deactivation associated to the emotion (Pekrun, 2019; Russell, 2003). Emotions can be grouped according to their combination of valence and arousal, with each group having a different relationship to performance (Harley et al., 2019a; Pekrun, 2019). Positive-activating emotions (like joy) facilitate learning and correlate to more accurate performance by associating to more flexible thinking and motivation towards the task (Pekrun & Perry, 2014). Negative-deactivating emotions, like sadness and boredom, are less beneficial for success because they cause disengagement, and lead to the avoidance of similar future situations due to the displeasing sensation (Pekrun & Perry, 2014). Negative-activating and positive-deactivating have a less clear connection to performance. Negative-activating emotions (i.e., anxiety, anger, confusion) might cause displeasing feelings towards the task, yet the physiological activation may trigger the person to act and seek solutions (Pekrun & Perry, 2014). Conversely, positive-deactivating emotions, like relief, lead learners to invest less energy on the task, yet the pleasant feeling will motivate learners to re-engage in similar future tasks (Pekrun & Perry, 2014).

We recognize the inherent co-occurrence of emotions and cognition, as shown in our guiding theory (Pekrun, 2019); however, we emphasize that this manuscript is focused on emotions. We see emotions as autonomous reactions that guide instinctive actions (Damasio, 2005). In the context of flight training, as pilots are highly concentrated in the task, we use autonomous physiological and behavioral reactions, namely emotions, to infer quick decision making (Damasio, 2005).

Emotions' Dynamics

Emotions are dynamics processes (D'Mello & Grasser, 2012; Zheng et al., 2023). Emotions are conventionally studied according to their tendency like categories (e.g., joy, boredom) and duration (S. Li, et al., 2021a). When learning a new skill, emotions might change constantly as the task evolves (Martins, 2016; Pekrun, 2019). Recent advances in learning sciences suggest the need to account for emotions' dynamics, defined as the patterns and regularities of fluctuations of emotions over time (Houben et al., 2015; Kuppens, 2015). Dynamic features of emotions are observed in frequency, intensity, and variability (Bailen et al., 2019; Krone et al., 2018; Zheng et al., 2023a). The frequency of emotions reflects how many times an emotion is experienced in a designated period of time (Bailen et al., 2019). Emotional intensity shows "the strength and magnitude of emotional response" (Bailen et al., 2019, pp. 64). Emotional variability emphasizes the fluctuations among multiple emotions over time (Thompson et al., 2012; Trull et al., 2008). Notably, there emotional features are complementary and can overlap. For example, a participant can answer a question reporting that they experienced a positive emotion scoring on 4 out of 10 strengths, showing that they experienced an emotion (having a frequency of one) but the intensity was low. Similarly, emotional variability can account for fluctuations of multiple emotions as they occur (i.e., frequency) or filtering only emotions with an intense presence (Zheng et al., 2023a). Emotion dynamics are a becoming a phenomenon of interest in learning and educational context as technology to detect nuanced and quick changes of emotions advances (D'Mello & Grasser, 2014; Duffy et al., 2016).

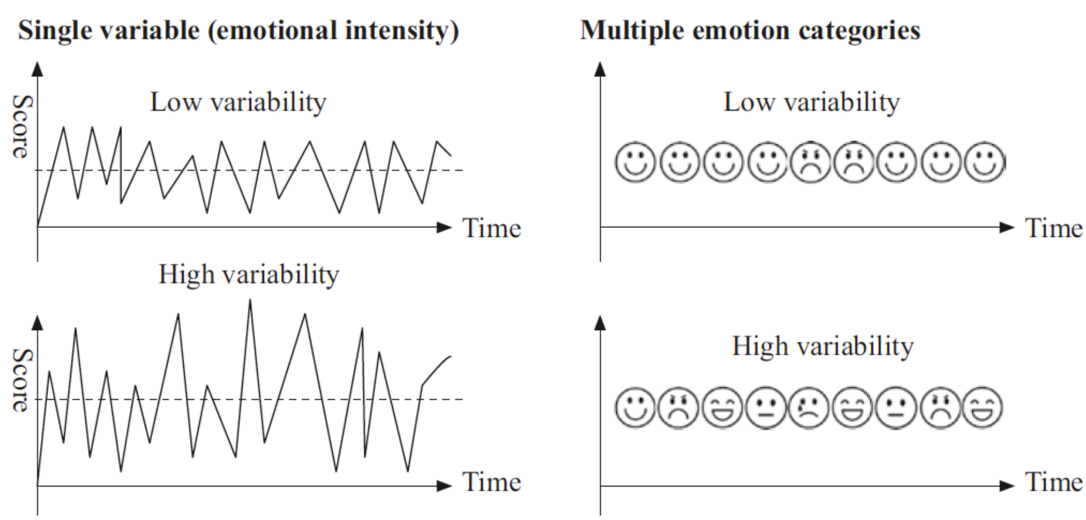
For this manuscript, to avoid confusion, we use arousal and intensity as different terms. Arousal refers solely to the physiological changes associated to emotions, namely, physiological activation denotes increases in arousal, whereas physiological deactivation shows decrease in arousal (Pekrun, 2019). Commonly, arousal is interpreted as physiological intensity, for instance outstanding peaks of physiological arousal (i.e., skin conductance responses) (Harley et al., 2019a). In this manuscript, we will also refer to intensity of behavioral changes, specifically degree of presence of facial expression of emotions. In conclusion, emotional intensity is used as an umbrella term to detect changes in the degree of emotional responses that could come from different measures, such as self-reports, physiology, or behaviour.

Research exploring the use of flight simulations for training has mainly focused on the role detrimental impact of negative-activating affective states, like stress and anxiety (i.e., Allsop & Gray, 2014; Hart, 2006). In the case of research exploring learning, the focus is extended to include both negative and positive valence; however, emotions tend to be studied as discrete states, rather than pairing emotional fluctuations according to the task demands (Duffy et al., 2016; Pekrun & Perry, 2014). For that reason, in this manuscript we argue that emotions dynamics can visualize the functionality of emotions as an adaptive response that go beyond emotional valence (i.e., (un)pleasantness) (Kashdan & Rottenberg, 2010). Particularly, this study contributes to understanding emotion dynamics in the context of pilot training using continuous and non-invasive measurements. It is expected that the results of this study identify the patterns of beginner pilot trainees' emotion dynamics and their potential function in parallel to performance accuracy changes.

In this study we focus on frequency of intense emotional reactions and emotional variability. In particular, the study focuses on emotional intensity (single-variable emotional responses), by exploring peaks of physiological arousal and changes in discrete emotions inferred from facial expressions. Emotional variability is analyzed using fluctuations of multiple facial expressions (Figure 3). This study focuses on authentic reactions during the learning process, and emotions are not artificially induced.

Figure 3

Illustration of Emotional Dynamics



Note. From "A review of measurements and techniques to study emotion dynamics in learning" by J. Zheng, S. Li, and S. Lajoie, in V. Kovanovic, R. Azevedo., D.C. Gibson, and D. Ifenthaler (Eds), *Unobtrusive observations of learning in digital environments.* (p. 10), 2023, Springer. Copyright 2023 Springer. Reproduced with permission.

This study is guided by a multimodal approach by combining physiological and behavioral responses to infer single-variable emotions (Han et al., 2020; Harley et al., 2015). Emotional intensity (strength of single-variable emotional responses, Zheng et al., 2023) will be inferred from the skin conductance responses (SCR) of electrodermal activity (EDA) and single dominant facial expressions. Emotional variability will be inferred from fluctuations among multiple dominant emotions.

To inform our understanding of emotional changes across training phases and difficulty levels, the section below reviews and discusses findings of previous studies exploring affective changes and performance as trainees go through flying simulations.

Emotional Changes as Flight Progresses and according to Difficulty Levels

Other studies in flight training have been interested in understanding emotional as the flight evolves and comparing difficulty levels. However, to our knowledge, both dimensions (time and difficulty) have not been examined together in a same study. Additionally, performance is explored only on some occasions.

Previous studies have confirmed that emotions change as the flight progresses, having distinctive effects on flying performance. Anxiety can have contrasting effects on flying performance, it can associate to less optimal cognitive processing in secondary tasks, but anxiety might not impact performance significantly (Hidalgo-Munoz et al., 2018). Long flights related to participants reporting more negative-deactivating emotions, a decrease in positive-activating emotions, and an increase in heart rate variability (Rosa et al., 2021, 2022). However, positive emotions increase when pilots are provided with more opportunities to practice (Koglbauer et al., 2011). These findings showcase the need to evaluate emotions in context (Lajoie, 2021; Pekrun, 2019).

More studies reported increases in negative-activating emotions with increases in flight difficulty level. Studies exploring affect in flight training confirm that pilots and pilot trainees tend to have more displeasing and unstable affective experiences during difficult flying maneuvers (Skibniewski et al., 2015; Tichon et al., 2014). However, the relationship between displeasing affect and flying performance has only been hypothesized to impact consequent performance, since studies had relied on pilots' perceptions of stereotypical scenarios, without assessing an actual flight (Hart & Bortolussi, 1984).

In the context of flight training, only Gaetan et al. (2015) discussed the implications of emotional variability in a descriptive manner; however, to our knowledge, pilots' emotional variability has not been previously quantified. Gaetan et al. (2015) examined the interaction between emotions and difficulty level during simulated flying tasks by describing patterns of participants according to similarity to expert performance. They found that the physiological arousal (i.e., muscle tension and electrodermal activity) of expert-like pilots increased with difficulty levels, and these participants reported experiencing more positive emotions with low intensity tasks. On the contrary, novice-like participants had increases in physiological arousal aligned with perceived workload, and they reported experiencing positive and negative emotions with high intensity and high variability. Intermediate pilots were like novices in presenting increases in physiological arousal aligned with perceived workload but reported more positive emotions with low intensity tasks, similar to expert reports. Although the authors did not report how they measured expertise, we believe that they pose an interesting new dimension to understanding the experience of pilots, that can account for emotional variability.

Research in medical diagnostics has investigated the impact of emotional variability in the process of diagnosing virtual patients (S. Li, et al., 2021a; S. Li, et al., 2021b). Findings show that medical students experienced more emotional variability during high-difficulty tasks as compared to low-difficulty ones, and high performers, who reached a correct diagnosis, experienced lower emotional variability, independently of task difficulty (S. Li, et al., 2021b). Furthermore, better diagnostic performance, similar to an expert solution, negatively correlated to medical students' emotional variability (S. Li, et al. 2021b), overlapping with Gaetan et al.'s (2015) findings in which more novice pilots had more emotional variability.

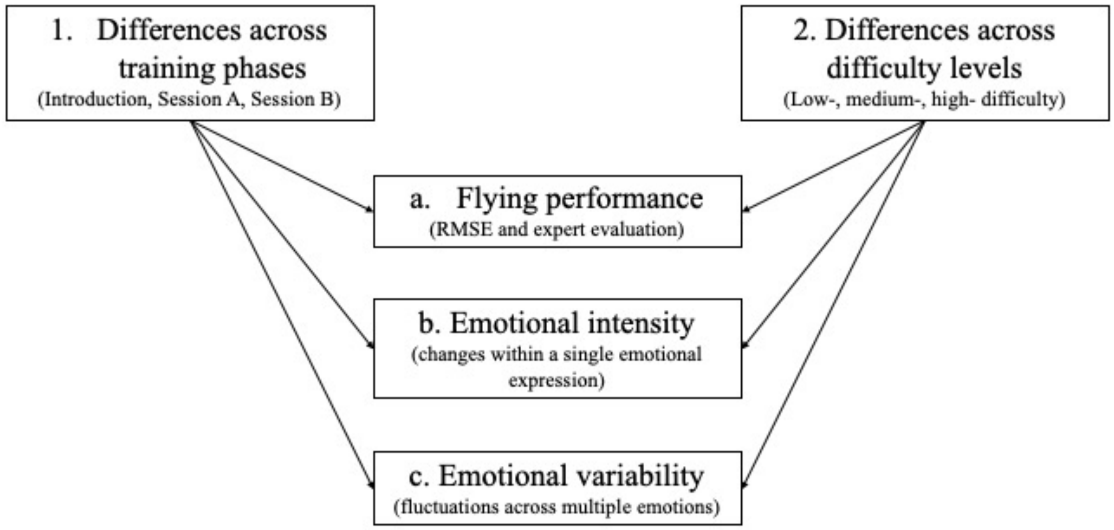
Current Study

This study attempts to examine the evolution of pilot trainees' performance and emotion dynamics according to training phase and difficulty level in a simulated flying task. We expect to contribute to the literature by exploring flying performance relative to behavioral and physiological emotional dynamic changes across time and difficulty level. Previous aviation research has explored emotional dynamics focusing on unique emotional expressions, such as intensity inferred from frequency of emotions and physiological responses (T. Li & Lajoie, 2021; Zheng et al., 2023a). To our knowledge, this would be the first study in aviation to quantitatively explore fluctuations among multiple emotions (i.e., emotional variability). We pose the following research questions and hypotheses. Figure 4 shows a diagram summarizing our approach.

1. How did (a) flying performance, (b) intense single expressions of emotions, inferred from behavioral (facial expression) and physiological responses (SCR); and (c) emotion variability (inferred from facial expression) differ across training phases?
2. How did (a) flying performance; (b) intense single expressions of emotions, inferred from behavioral (facial expression) and physiological responses (SCR); and (c) emotional variability (inferred from facial expression) differ across difficulty levels?

Figure 4

Visual Representation of Hypotheses and Research Questions



Hypotheses

1) Across training phases: Training phases were consecutive and divided as introduction (feedback provided), sessions A and B (independent execution without feedback).
    a. Performance and emotions will vary across training phases. Performance will be better in the final practice (session B), as compared to introduction and session A.
    b. Positive emotions (i.e., happy) will be more frequent during introduction compared to session A and session B (Rosa et al., 2021, 2022) since participants might be more energetic and motivated at the beginning of the task. Negative-activating emotions (i.e., anger, disgust, and fear) and SCR will be more frequent in introduction, since participants are expected to be more engaged at the beginning of the task, and session A because participants are expected to have more difficulties solving the tasks independently (with no feedback), compared to session B, when participants will have had sufficient practice and might be more tired (Hidalgo-Munoz et al., 2018; Koglbauer et al., 2011). Negative-deactivating

emotion (i.e., sad/ boredom) will be more frequent in session B compared to introduction and practice A because participants might be more tired at the end of the task (Rosa et al., 2021, 2022). Surprise is known to occur after an unexpected event while flying (Kinney & O'Hare, 2020), therefore, we expect surprise to be higher during introduction phase due to the novelty of the task (Landman et al., 2020).

c. Emotional variability (i.e., fluctuations among multiple emotions) will be higher during introduction and first session A, as we expect participant to be less emotionally engaged in session B. Practice and familiarity might imply that the task is less stimulating, causing less emotional fluctuations (Rosa et al., 2022).

2) Across difficulty levels: Performance and emotions will vary across task difficulty.

a. Performance will be poorer in high difficulty, as compared to low and medium-difficulty levels (S. Li, et al., 2021b).

b. Positive emotions will be higher in low-difficulty as compared to medium and difficult levels (D'Mello & Graesser, 2012; Pekrun, 2019). Negative-activating emotions will be higher in difficult tasks, compared to low and medium levels (Hart & Bortolussi, 1984; Skibniewski et al., 2015; Tichon, Wallis, et al., 2014). Negative-deactivating emotion will be higher during low-difficulty tasks, compared to medium and difficult tasks since low-difficulty will require less cognitive engagement (D'Mello & Graesser, 2012). We expect that difficult tasks will cause more cognitive dissonance from expectations, compared to easy and medium difficulty-tasks, causing participants to express more surprise during difficult tasks (D'Mello & Graesser, 2012).

c. Emotional intensity (single-variable emotional reactions), inferred from skin conductance responses and, emotional variability (fluctuations among multiple emotions), will increase with difficulty level, being higher in high-difficult tasks compared to low and medium difficulty (Gaetan et al., 2015; S. Li, et al., 2021a; Skibniewski et al., 2015).

Methods

Participants

This study was part of a larger project with the objective of measuring cognitive and affective processes of ab-initio pilot training with a flight simulation. Only relevant methods are presented. Volunteers were recruited to mimic junior pilot trainees with little to no experience flying airplanes (Marques et al., 2023). The only requirements for applying to become a student pilot in North America is being 14 years of age or older, being able to read, write, speak, and understand English, and passing a medical evaluation (FAA Department of transportation, 2003; Transport Canada, 2019). These same requirements were used to recruit participants for this project. The larger project included multiple physiological sensors (i.e., EEG, heart rate, EDA) and participating in a two-day data collection. The overall data collection required approximately eight hours, which included travel time to the location. Due to the length of the study and the fact that it was conducted during the pandemic recruitment was convenience based, targeting potential volunteers who had the time flexibility and willingness to participate with awareness of the experiment demands.

Following ethics approval, twenty-three volunteers (M_{age}=28.96, SD=4.68) were recruited from a large North American city; 12 self-identified as females (52.2%) and 11 as males (47.8%). Participants had diverse educational backgrounds, including high school degree currently studying to obtain college and a bachelor's degree (*n*=2), bachelors (*n*=9), and masters (*n*=12) degrees. The participants who had University degrees were mostly from STEM fields (*n*=11), followed by studies in finance and accounting (*n*=3), psychology-related domains (*n*=3), nutrition (*n*=2), and neuroscience (*n*=1). One participant graduated from aviation school obtaining an US Federal Aviation Agency (FAA) commercial pilot rating, the prerequisite to obtaining a commercial pilot certificate (FAA Department of transportation, 2003) with experience using Microsoft Flight Sim and X-Plane. Other five participants reported having experience using flight simulations four of them had used a flight simulation one time, and one did not specify. The facial expression of the participant with the commercial pilot rating was not recorded, thus, not included in the analyses using facial expression. Participants who were screened into the study reported to be free from any medical condition that could limit their participation. Participants signed a consent form informing them of the purpose of the study, emphasizing voluntary participation.

G*power software application was used to conduct a power analysis a priori to define the sample size for conducting within subjects repeated measures ANOVAs. For comparing three groups (i.e., three training phases or three difficulty levels) and 13 measures (i.e., four measures of performance, seven emotions, one measure of skin conductance, and one measure of emotional variability), and their corresponding post-hoc analyses, revealed that a sample size of 12 participants is needed to achieve a power of 0.80, assuming a significance level of 0.05 and a medium effect size of 0.25 (Cohen, 1988; Faul et al., 2007). Using the same statistical parameters for comparing the changes in facial expression of the seven emotions across each training phase and each difficulty level, G*power analysis revealed that a sample of 17 participants would be need.

Apparatus

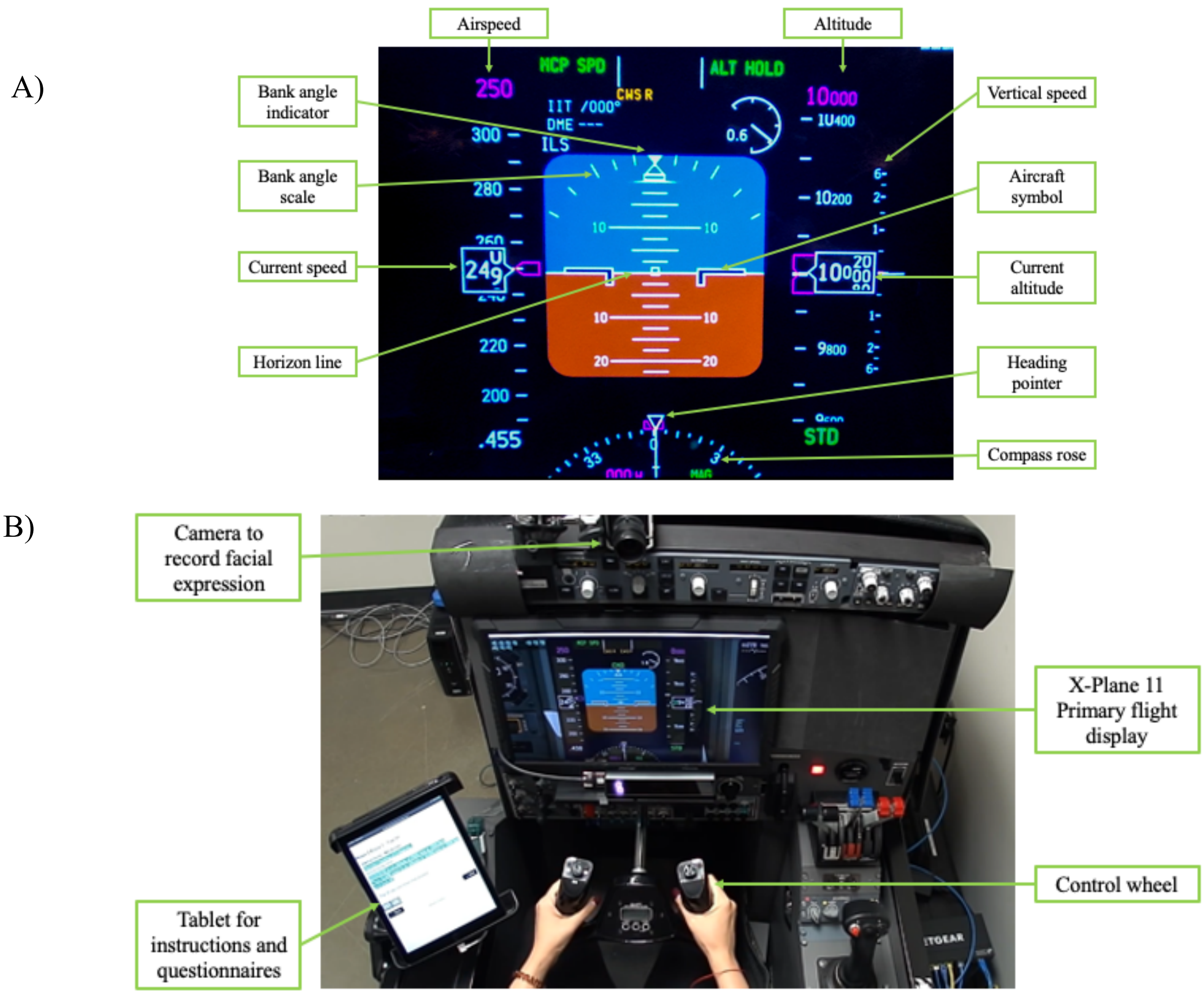
The experiment consisted of training ab initio pilots to perform flying twizzles (i.e., basic flying maneuvers) in a fixed-base simulator designed and operated by Marinvent Corporation. The cockpit included a control yoke, throttle, and pedals, and a screen which showed an aircraft primary flight display (see Figure 5). The throttle and pedals were operated by an autopilot. Participants used the yoke to control aircraft roll, by turning the yoke left and right, and aircraft pitch, by moving the yoke forward and backwards. Aircraft bank angle primarily affects aircraft turn rate and is used to control heading, while the aircraft pitch angle is used to set climb/descent rates and thus control altitude. The primary flight display was rendered using X-Plane 11, which is a flight simulation software package designed to reflect the behaviour of real aircraft (Figure 6-A, Laminar Research, 2022). The investigating team installed a camera on top of the primary flight display to record participants' facial expression. A tablet was set up on the left side of the control wheel for participants to read instructions and answer questionnaires (see Figure 6-B).

Figure 5

Flying Simulator Cockpit

Figure 6

X-Plane 11 Primary Flight Display



Flying Tasks

The twizzles were designed by an expert pilot instructor. Participants were required to perform changes in direction (i.e., turns), and altitude (climbs and descents) while only referring to instrument indications with no outside visual cues. Flight maneuvers are dynamic processes, where changes in one metric will affect another. For example, increasing altitude will decrease airspeed. Therefore, in this experiment airspeed was set to autopilot programmed to maintain a steady 250 knots.

The flying maneuvers had three difficulty levels: low, medium, and high. Low difficulty maneuvers required a change in one axis (roll or pitch) to achieve a target (heading or altitude) and then maintenance of the new heading or altitude. Medium difficulty maneuvers required a change in one axis (heading or altitude) while maintaining the other axis, followed by a reversal to the original flight condition (i.e., starting heading and altitude), and maintenance of the original conditions. High difficulty maneuvers required simultaneous changes in two axes: a change with reversal in one axis and a change without reversal in the second axis. See Table 3 for example maneuver instructions.

Table 3

Sample Instructions for Maneuvers According to Difficulty Levels

Difficulty level	Instructions
Low	Maintain altitude at 10,000 feet. At the same time: Turn LEFT at 30 deg AOB to a heading of 240 degrees and roll out on a steady heading.
Medium	Maintain altitude at 10,000 feet. At the same time: Turn LEFT at 30 deg AOB to a heading of 300 degrees; Then turn RIGHT at 30 deg AOB back to a heading of 0 degrees and roll out on a steady heading.
High	DESCEND to an altitude of 9,000 feet at 1000 fpm and level off. At the same time: Turn LEFT at 30 deg AOB to a heading of 300 degrees; Then turn RIGHT at 30 deg AOB back to a heading of 0 degrees and roll out on a steady heading.

Procedures

Upon arrival, the goals of the experiment were explained to the participants before they signed the consent form. Participants filled out questionnaires regarding demographic information and relevant previous experience. Participants received instructions on how to use the cockpit and perform flying maneuvers, emphasizing aviation-related language. Afterwards, researchers set up electrodes to detect EDA. The experiment was divided into three main phases: introduction (seven trials), session A (eight trials), and session B (seven trials), for a total of 22 trials. Each trial had a 30 second baseline (“maintain straight and level” at 10 000 feet altitude, and a heading of 0°), followed by a 90 second twizzle maneuver. The X-plane 11 instrument display and aircraft state were re-set before every trial: participants would start from 10,000 feet altitude, 250kts (automatically controlled) airspeed, 0° bank angle and 0° heading. The introduction phase was the same for all participants, with a sequence of increasing difficulty: four low-difficulty, two medium-difficulty, and one high-difficulty maneuvers. During the introduction phase participants received feedback from a trained researcher to confirm that participants understood the instructions and how to use the control wheel and displays to complete the task. During sessions A and B, the task difficulty order was randomized, and participants completed five trials for each difficulty level. During sessions A and B, participants no longer received feedback from a researcher. Participants could read the twizzle maneuver instructions at their own pace and start the task when they were ready. Participants were allowed to take breaks between each phase. The full procedure, see Table 4, took around 6 hours.

Table 4

Sample Procedures of Experiment

Sign consent	
Demographics	
Video-training	
Guided hands-on familiarization	
Training phase	
1.	Low
2.	Low
3.	Low
4.	Low
5.	Medium
6.	Medium
7.	High

First practice	
8.	Medium
9.	Low
10.	High
11.	Low
12.	High
13.	Medium
14.	Low
15.	High

Second practice	
16.	Medium
17.	Low
18.	High
19.	Medium
20.	High
21.	Medium
22.	Low

Measures

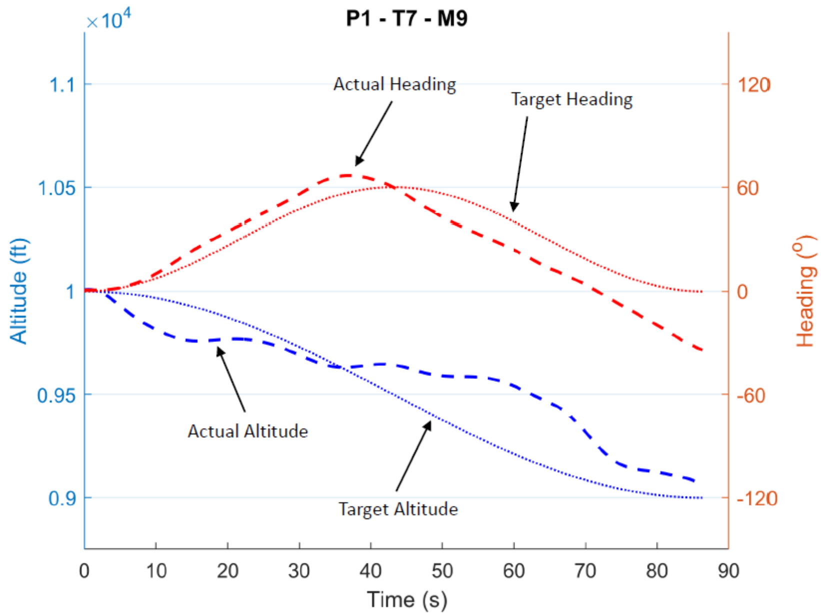
Flying Performance

Flying performance was assessed from aircraft state data contained in the X-Plane 11. The metrics of altitude and heading were selected to evaluate task performance in pitch and roll axes. Performance was measured using two metrics: 1) flying error and 2) expert rating.

Flying error. Root-mean square error (RMSE) is the most common measure to assess flying performance in aviation (Allsop & Gray, 2014; Gray et al., 2016). RMSE is calculated by subtracting actual altitude or heading from the target altitude or heading of each time point, squaring the error, and computing the square root of mean error. Since the tasks in this experiment are continuous and, on some occasions require reversal paths, sinusoidal target functions were used to connect the starting, middle, and end points for instructed heading and altitude across the 90 seconds allocated to each twizzle (Jennings et al., 2024). Thus, the target metric accounted for expected continuous changes for an idealized flight path, see Figure 7 for a visual representation. RMSE reflects error regarding the deviation between ideal flight and actual performance: a larger number reflects more deviation from ideal flight, thus more error.

Figure 7

Example of RMSE Calculation

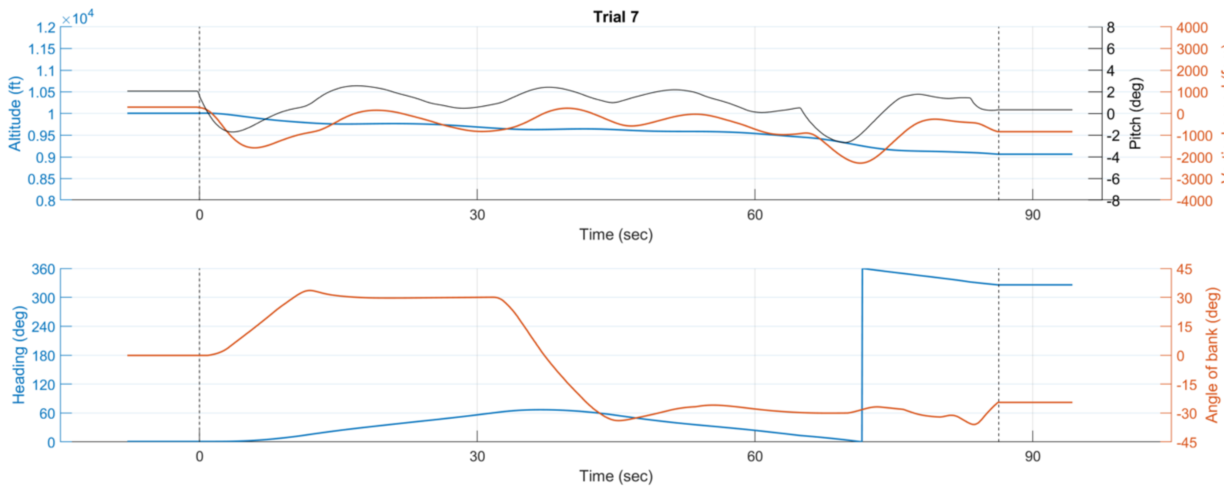


Note. The instruction of this example was as follows: "Descend to an altitude of 9,000 feet at 1000 fpm and level off. At the same time: Turn right at 30 degrees angle of bank to a heading of 60 degrees. Then turn left at 30 degrees angle of bank back to a heading of 0 degrees and roll out on a steady heading."

Expert rating. Expert aviation instructors rated participants performance after observing graphs with their flying trajectories (see Figure 8). Scores ranged from one to four, one represented very low accuracy compared to the instruction, and four would represent high accuracy compared the target metric. Zero was set when the task was not carried out. Consequently, a higher score would represent higher performance. This kind of scoring is commonly used for rating trainees' performance by the industrial partner in this project. The scoring is usually done with the instructor observing trainees' performance in the simulation. However, this study uses an innovative approach whereby instructors would examine and score graphs that summarized the individual participant's flying path. Preliminary results of this new approach (evaluating flying path in graphs) demonstrate alignment with state data of the aircraft, and more studies are being performed to confirm its reliability (Jennings et al., 2024).

Figure 8

Sample Graphs for Expert Analysis



Note. The instruction of this example was as follows: "Descend to an altitude of 9,000 feet at 1000 fpm and level off. At the same time: Turn right at 30 degrees angle of bank to a heading of 60 degrees. Then turn left at 30 degrees angle of bank back to a heading of 0 degrees and roll out on a steady heading." Expert rated altitude in a score of 4 and heading with a score of 3.

Emotional Intensity Inferred Through Skin Conductance Responses

Electrodermal activity shows changes in the autonomous nervous system from changes in skin conductance (Braithwaite et al., 2015). The main features of EDA include skin conductance levels (SCL) and skin conductance responses (SCR). SCL shows continuous changes of skin conductance and general autonomic arousal (Braithwaite et al., 2015). SCR shows rapid and intense changes (significantly higher than the individual's baseline) in physiological arousal, creating peaks in skin conductance (Braithwaite et al., 2015). This study focuses on SCR as a meaningful indicator of physiological arousal to infer emotional intensity, and engagement towards learning and solving the task (Hardy et al., 2013). SCL, as a general autonomic response, might not provide additional information about emotional and cognitive engagement in the task (Boucsein et al., 2012).

EDA was recorded using BioSemi. Electrodes were positioned on the right hand, one at the hypothermal region of the hand and the second at the wrist, to detect changes in sweat to infer skin conductivity (see Figure 9). EDA was processed using NeuroKit2, an open-source python toolbox designed for neurophysiological signals processing (Makowski et al., 2021). In the case of EDA, raw data is used as the input, and NeuroKit2 returns the filter signal, phasic components, SCR onsets, indexes, and amplitudes (Makowski et al., 2021). NeuroKit2 includes the convex optimisation approach (cvxEDA) accounting for white Gaussian noise for controlling for artifacts and errors (Greco et al., 2016). Since the raw data is used, a personalized baseline is automatically created for each participant and filters out components of EDA, including phasic components, and SCR onsets, indexes, and amplitudes (Makowski et al., 2021). We concentrated on SCRs since they are significantly higher peaks that can be interpreted as rapid and meaningful changes resulting from psychological engagement, and thus interpreted a measure of emotional intensity (Boucsein et al., 2012; Hardey et al., 2013; Harley et al., 2019b). See Figure 10 for a visual representation of NeuroKit2 analysis and extraction of SCR.

Figure 9

Sample Location of Electrodes to Detect Electrodermal Activity.

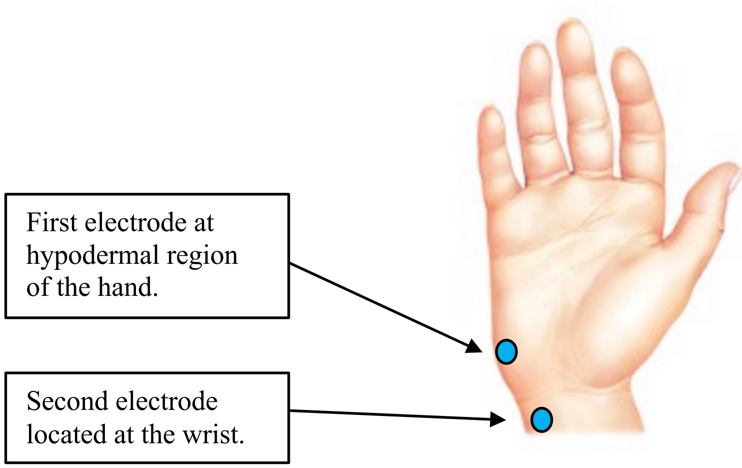
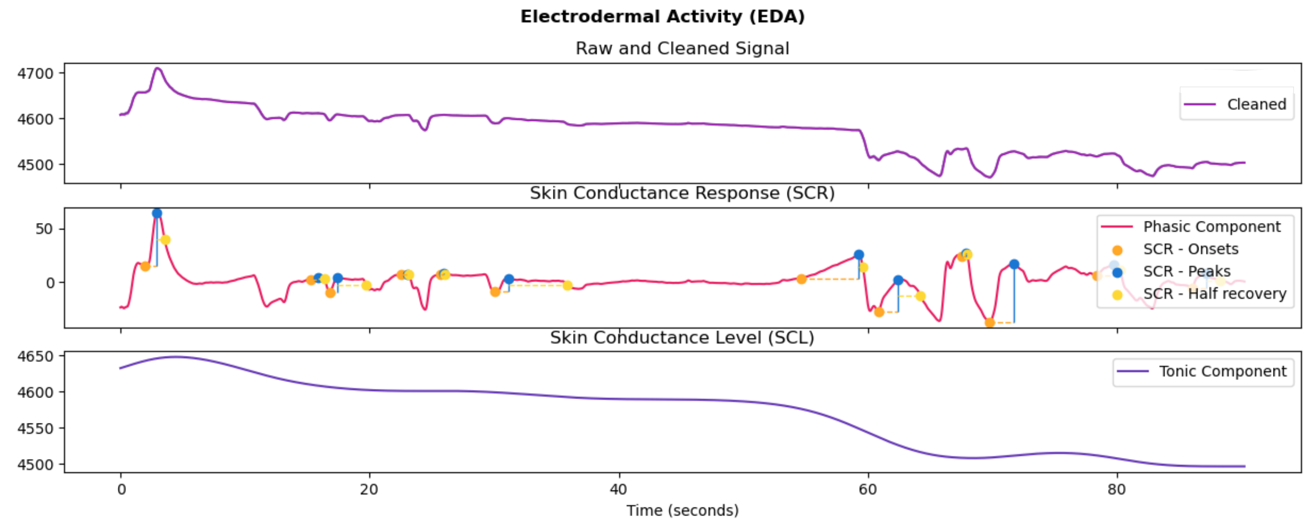


Figure 10

Sample of Clean EDA Signal



Note. The trial of this example was task two during introduction, with low difficulty. The instruction was as follows: "Maintain altitude at 10,000 feet. At the same time: Turn right at 30 deg AOB to a heading of 120 degrees and roll out in a steady heading."

Emotional Intensity inferred through Facial Expression of Emotions

Facial expressions of participants were recorded throughout the experiment using a camera mounted on top of the screen showing the primary flight display. Participants who used corrective lenses were required to use contact lenses to better identify facial expression. For identifying emotions, videos were processed using FaceReader 6.0 (Loijens et al., 2015). FaceReader is a software application trained to detect and analyse facial expressions (Loijens & Krips, 2021). FaceReader classifies emotions based in three main stages (Loijens & Krips, 2021). First, FaceReader detects the face using the Viola-Jones algorithm, a deep-learning algorithm (Viola & Jones, 2004; Zafeiriou et al., 2015). In the second stage, FaceReader uses an active appearance model, based on deep neural networks, to map and record 500 key points of the face (Cootes & Taylor, 2001; Loijens et al., 2015). Such key points are later merged using Principal Component Analysis to calculate a vector comparing the dimensionality of the face with a model (Loijens et al., 2015). In the last stage, the vector is used as the input in an artificial neural network (ANN) to identify the emotions classification (Loijens et al., 2015). FaceReader's ANN was trained using 10,000 manually annotated images following the Facial Action Coding system (Ekman & Rosenberg, 2005). The ANN is trained to classify facial expression of seven basic emotions: anger, disgust, happy, neutral, sad, scared, and surprised with an accuracy of 90% (Ekman, 1992; Loijens et al., 2015). Due to individual differences in facial expression, FaceReader was calibrated for each participant by selecting a representative neutral facial expression before performing the analysis.

The sample rate used for FaceReader was 30 samples per second (Loijens et al., 2015). For each time point, the intensity of each emotion is recorded in a scale of 0 to 1, in which 0 indicates that the emotion is not present, and 1 implies that the emotion is fully present (Loijens et al., 2015). For the current study, we used the frequency of dominant emotions, using the state log output. Dominant emotions are recorded in the log if they meet the following criteria: (1) emotions are sustained for more than 0.5 seconds, and (2) emotions have a higher intensity (or presence) compared to the other emotions (Loijens et al., 2015). For that reason, we argue that the dominant emotions output represents intense and meaningful emotions.

The frequency of dominant emotions was counted per training phase and per difficulty level. Such frequency was used in a first analysis to identify difference among type of emotions expressed across training phases and difficulty levels. Additionally, the frequency of emotions was used in a second analysis to calculate emotional variability.

Notably, our guiding theory is the control-value theory of emotions (Pekrun, 2019); however, FaceReader, categorizes emotions based on Ekman's basic emotions theory (1992). Therefore, we rely on Harley et al. (2015) empirical pairing of facial expression of achievement and basic emotions to interpret our results: frustration is aligned with anger, confusion with disgust, joy with happiness, anxiety with fear, and boredom with sadness. Surprise cannot be paired directly as an achievement emotion; however it is included for its relevance in flight training and performance (Landman et al., 2017). Although the definition of neutral as an emotion is debatable (Lajoie et al., 2021; Russell, 2003), this study will follow Harley et al.'s (2012) proposal of accounting neutral as a baseline state, indicating that trainees are not emotionally distracted and can learn and, neutral will facilitate the identification of fluctuation of emotions.

Emotional Variability Using Facial Expressions

Emotion variability refers to the fluctuations in emotional states and provides a dynamic insight into emotions above and beyond the frequency of emotions. In this manuscript, we calculate the variability, or degree of randomness of different occurring emotions, of intense (i.e., dominant) emotions using Shannon's (1948) entropy formula (Jack et al., 2014; S. Li, et al., 2021a):

$$h(p_1,\ldots,p_a) = -\sum_{i=1}^{a} pi\log_2(p_i)$$

In which, p_i is the probability of emotion *i* occurring in a sequence of emotions. In this case, the sequence includes the expression of the seven basic emotions. Using the binary logarithm ($\log_2$) helps us identify the times the number needs to be multiplied by itself to obtain pi, therefore, the sum of the $\log_2$ of the probabilities of the seven categories of emotions would show the degree of randomness (Karaca & Moonis, 2022). Specifically, the minimal value of entropy is zero, implying that the person expressed the same emotion throughout the task, with only one emotion having the highest probability of occurring. The maximum value of entropy was 2.8 ($\log_2(7)$), implying that the person experienced the seven emotions with equal probability, thus having high emotional variability (Zheng et al., 2023a).

Notably, zero does not equal neutral, but rather shows persistence of only one of the seven emotions. Neutral is treated as a dominant emotion, and its frequency was transformed to calculate its probability of occurring. Thus, having neutral in the dominant emotions log would imply that this expression was sustained for more than 0.5 seconds and that its presence was higher than the other categories (Loijens et al., 2015).

Results

Data Screening

Within-subjects repeated-measures ANOVAs (RM-ANOVAs) were conducted in IBM® SPSS® version 29 to address each of our research questions. Python's package Matplotlib was used to create graphs (Hunter, 2007). Assumptions of independence and sphericity were met for all variables. Normality assumption was assessed with Shapiro-Wilk test (Meyers et al., 2016), showing a violation of normality for facial expression of disgust, fear, happy, sad, and surprise across the three training phases and three difficulty levels. RM-ANOVAs were conducted to detect variables with statistical significant differences, additionally, pairwise comparisons were conducted to evaluate the differences among the individual scores of the variables, using the Bonferroni method to control for Type I error, in which the expected alpha value to identify significance (i.e., p=.05) was divided by the number of repeated measures (i.e., seven emotions, three training phases, or three difficulty levels) for a cut-off value of p=.007 and p=.017 accordingly.

Two clarifications are provided here regarding the data used for analysis. First, during the experimental process, the facial expression of the participant with FAA instruction rating could not be analyzed because the participant wore a face mask throughout the experiment, thus analyzes regarding frequency of emotions and emotional variability were performed only with 22 participants and flying performance and SCR analyses were conducted with 23 participants. Secondly, analyses exploring differences across difficulty levels only included trials in sessions A and B, such that there was an equal number of trials (i.e., five) per difficulty level.

Due to the heterogeneity of the participants, non-parametric statistical analyses were conducted to identify if the distribution of the variables differed according to participants' backgrounds. Independent-samples Kruskal-Wallis' tests were conducted to identify differences across educational background and Mann-Whitney U tests were conducted to identify differences among participants with and without previous experience using flight simulators, for statistics consult Appendix A. Results showed that flying performance and emotional reaction of participants with a bachelor's and a master's degree differed when comparing educational background in three cases: participants with a bachelor's degree expressed more anger during low-difficulty tasks (p=.036), and had less neutral facial expressions during high-difficulty tasks (p=.019); and, participants with a master's degree had a higher altitude RMSE during medium difficulty tasks (p=.026). Flying performance and emotional reactions did not differ regardless of previous experience with flight simulators.

Differences Across Training Phases

Flying Performance

Training phase had a significant effect on altitude performance measures, and it did not have a significant effect on heading metrics, see Table 5. Altitude RMSE was significantly higher in the introduction phase compared to session B (p<.001), and it was marginally higher in introduction compared to session A (p=.024) according to Pairwise comparisons. Expert altitude scorings also increased across phases: the introduction phase had a lower score than session A (p=.006) and session B (p<.001); and session A had a lower score than session B (p=.006). Overall, altitude error was lower at the end of the experiment (session B), compared to the introduction phase. See trends in Figure 11.

Table 5

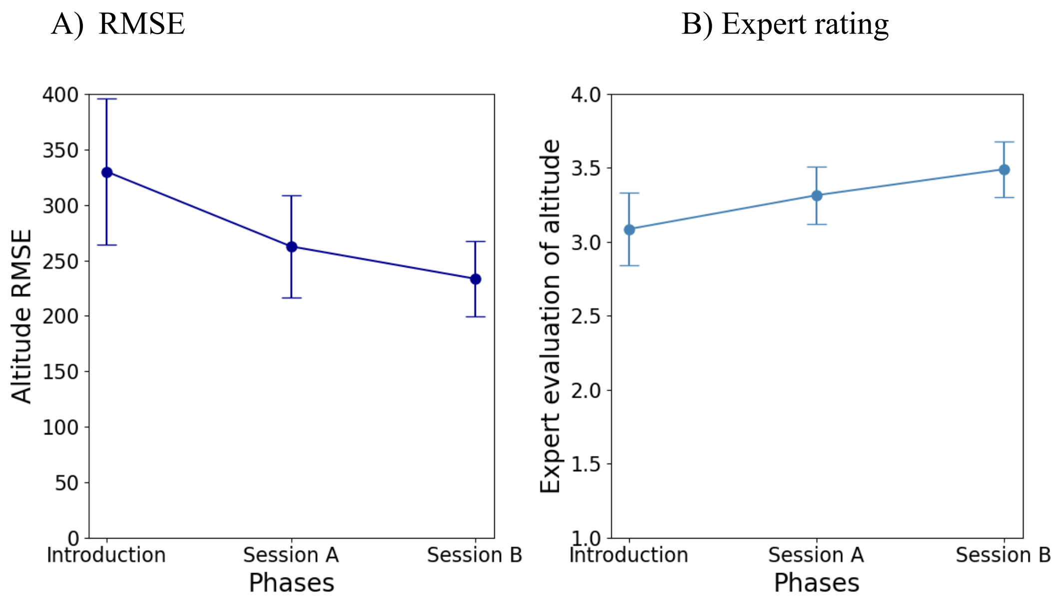
Flying Performance Across Training Phases

Performance measure	Practice phase	Mean	SD	MD	F	p	η^2
Heading RMSE	Introduction	10.12	4.64		.079	.930	.003
	Session A	10.21	3.33				
	Session B	10.48	3.81				
Altitude RMSE	Introduction	329.98	161.97	I>SA*	9.566	.002	.303
	Session A	262.57	113.01				
	Session B	233.42	83.20	I>SB**			
Heading expert	Introduction	3.67	.33		1.077	.350	.047
	Session A	3.73	.36				
	Session B	3.76	.33				
Altitude expert	Introduction	3.09	.60	I<SA**	15.79	<.001	.418
	Session A	3.31	.48	SA<SB**			
	Session B	3.49	.46	I<SB**			

Note. SD=Standard Deviation, MD=Mean Difference, I=Introduction, SA=Session A, SB=Session B. *p=.017 after Bonferroni correction, **p<.001.

Figure 11

Altitude Performance Across Training Phases



Note. Error bars represent confidence intervals.

Emotional Intensity - Physiological Arousal

Training phases had a significant effect on the frequency of SCR, descriptive statistics in Table 6. Pairwise comparisons showed that participants significantly experienced more SCRs during session A compared to introduction (p<.001) and session B (p=.011), see Figure 12.

Table 6

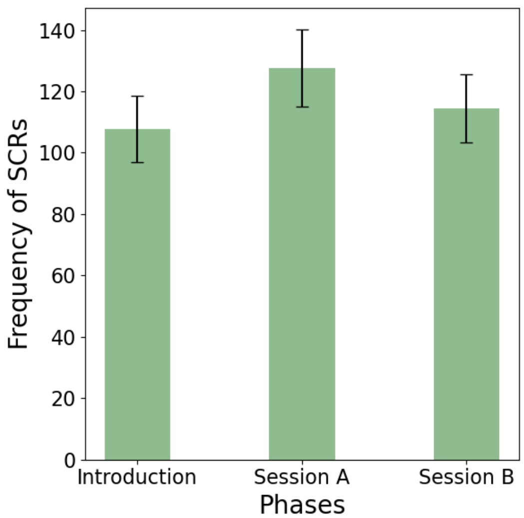
SCR across Training Phases

Performance measure	Training Phases	Mean	SD	MD	F	p	η^2
SCR	Introduction	107.69	25.57	I<SA**	7.98	.001	.266
	Session A	127.61	30.69	SA>SB*			
	Session B	114.43	27.44				

Note. SD=Standard Deviation, MD=Mean Difference, I=Introduction, SA=Session A, SB=Session B. *p<.017 after Bonferroni correction, **p<.001.

Figure 12

Skin Conductance Responses Across Training Phases



Note. Error bars represent confidence intervals.

Emotional Intensity - Facial Expressions

Three RM-ANOVAs were conducted, one for each phase (introduction, session A, and session B). In each RM-ANOVA, the frequency of the seven dominant emotions were compared (i.e., anger, disgust, happy, disgust, fear, sad, and surprise); descriptive statistics in Table 7. We used frequency of facial expressions, rather than proportions, since the duration of the tasks was equal for all participants (Lajoie et al, 2021). Frequency of facial expression of emotions had a statistically significant difference, see Figure 13. For the introduction phase (F=18.467, p<.001, η^2=.468), participants significantly had more expressions of neutral and anger compared to sad, happy, fear, and disgust expressions. A similar pattern was observed for session A (F=14.74, p<.001, η^2=.412) and session B (F=16.079, p<.001, η^2=.434), with the exception that the expression of anger did not have a statistically significant difference from sadness.

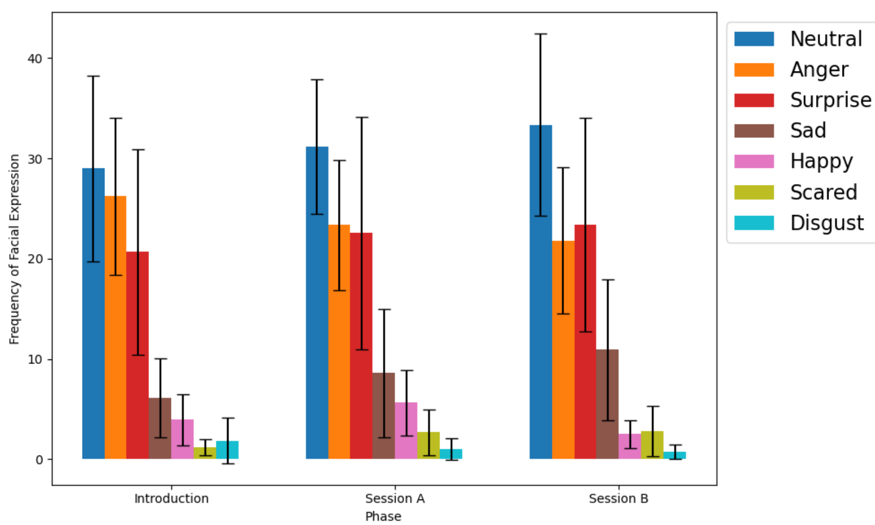
Table 7

Descriptive Statistics of Frequency of Facial Expressions Across Training Phases

	Introduction		Session A		Session B	
	M	*SD*	*M*	*SD*	*M*	*SD*
Neutral	29	22.18	31.18	16.09	33.36	21.27
Anger	26.23	18.71	23.36	15.53	21.82	17.52
Surprise	14.04	14.73	15.64	20.08	16.23	17.78
Sad	6.14	9.37	8.59	15.34	10.91	16.90
Happy	3.54	4.75	5.64	7.74	2.5	3.26
Scared	1.23	1.92	2.68	5.48	2.82	6.06
Disgusted	1	1.90	0.77	1.54	0.45	0.67

Figure 13

Frequency of Intense Facial Expressions across Training Phases



Note. Pairwise comparisons: NE>SA in introduction**, session A** and session B*. NE>HA, NE>SC, NE>DI in introduction**, session A** and session B**. AN>SA in introduction* and session A*, AN>HA, AN>SC, AN>DI in introduction**, session A**, and session B**. NE=Neutral, AN=Angry, SU=Surprise, SA=Sad, SC=Scared, DI=Disgust. *p<.007 after Bonferroni correction, **p<.001. Error bars represent confidence intervals.

As a follow-up, RM-ANOVAs were conducted to understand if single emotions differed across training phases. Results show a significant effect of training phases only on expression of sadness (F=3.868, p=.029, η^2=.156). Pairwise comparisons showed that sadness was significantly higher in session B compared to introduction ($t(1, 22)=-2.707$, $p=.013$, Cohen's $d=.57$).

Emotional Variability – Facial Expressions Fluctuations

Training phases did not have a significant effect on emotional variability. Descriptively participants experienced more emotional variability during session A, compared to introduction and session B (see Table 8 and Figure 14).

Table 8

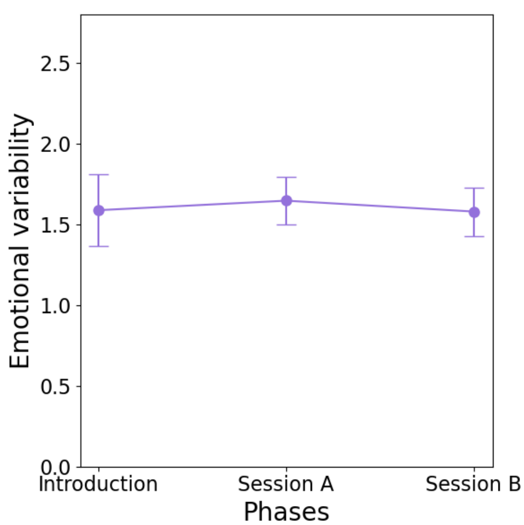
Emotional Variability Across Training Phases

Performance measure	Training phases	Mean	SD	MD	F	p	η^2
Emotional variability	Introduction	1.58	.55		.312	.734	.015
	Session A	1.65	.36				
	Session B	1.58	.37				

Note. SD=Standard Deviation, MD=Mean Difference, I=Introduction, SA=Session A, SB=Session B. *p=.017, after Bonferroni correction, **p<.001.

Figure 14

Emotional Variability Across Training Phases



Note. Error bars represent confidence intervals.

Differences Across Difficulty Levels

Flying Performance

Difficulty level had a significant effect on performance during session A and B; descriptive statistics in Table 9. Pairwise comparisons showed performance had statistically significant differences across difficulty levels.

Table 9

Flying Performance Across Difficulty Levels

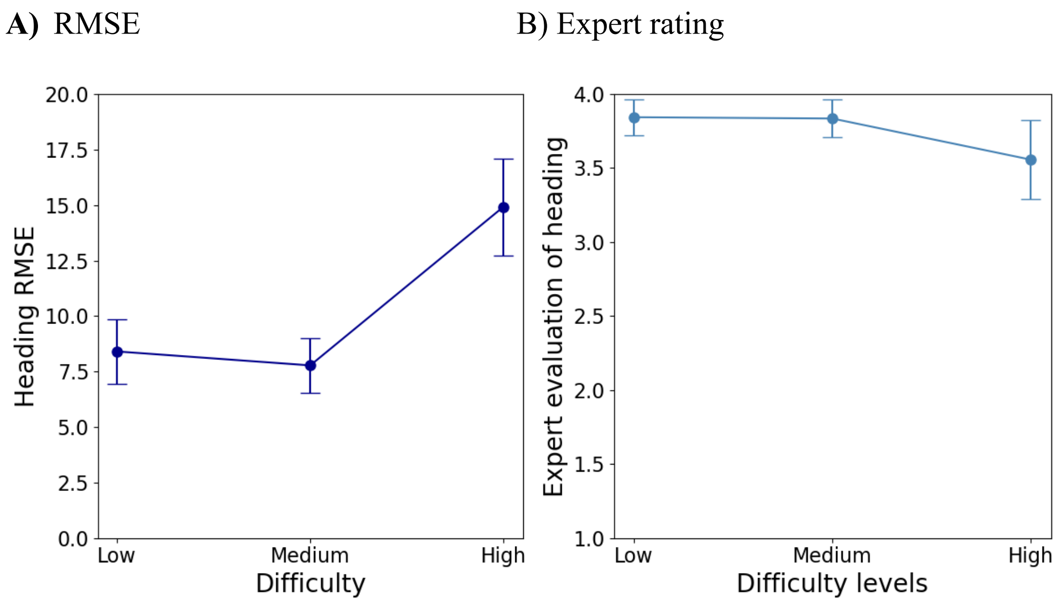
Performance measure	Difficulty level	Mean	SD	MD	*F*	*p*	η^2
Heading RMSE	Low	8.40	3.57	L<H**	31.11	<.001	.586
	Medium	7.77	3.02	M<H**			
	High	14.92	5.35				
Altitude RMSE	Low	186.55	81.65	L<H**	24.28	<.001	.525
	Medium	183.07	87.13	M<H**			
	High	377.54	188.45				
Heading expert	Low	3.84	.30		3.88	.028	.150
	Medium	3.83	.31	M>H			
	High	3.55	.65				
Altitude expert	Low	3.68	.41	L>H**	32.71	<.001	.598
	Medium	3.56	.47	M>H**			
	High	2.95	.65				

Note. SD=Standard Deviation, MD=Mean Difference, L=Low, M=Medium, H=High. *p=.017 after Bonferroni correction, **p<.001.

Heading RMSE was significantly higher for high-difficulty maneuvers, compared to low- (p<.001) and medium-difficulty maneuvers (p<.001). Participants had a lower heading score in high-difficult maneuvers compared to medium-difficulty, with a marginally significant difference (p=.02). See trends in Figure 15.

Figure 15

Heading Performance Across Difficulty Levels

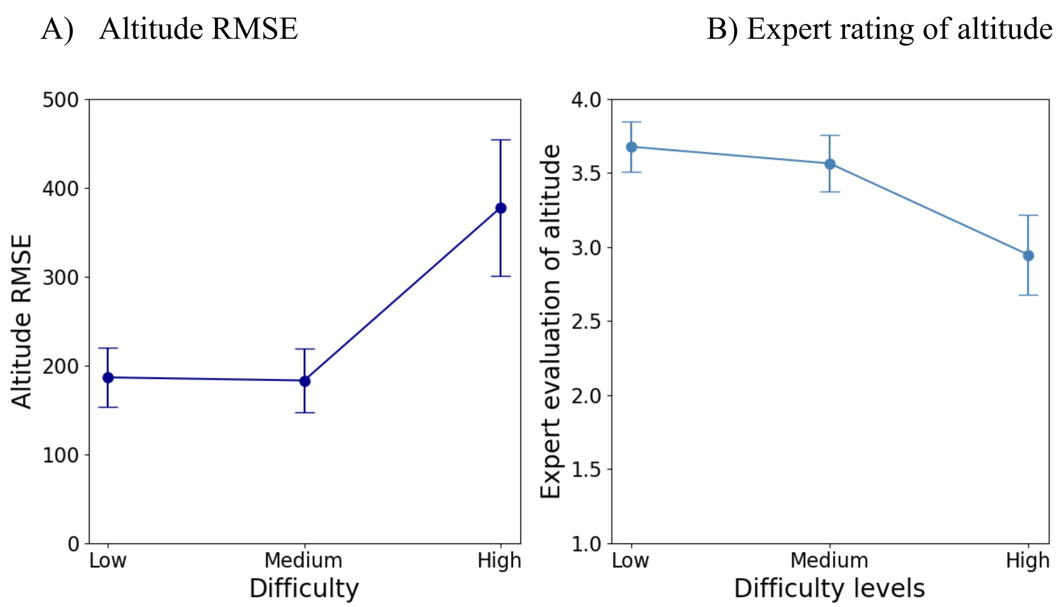


Note. Error bars represent confidence intervals.

Altitude RMSE was higher for high difficulty maneuvers compared to low (p<.001) and medium difficulty levels (p<.001). Participants had lower altitude scores for high difficulty maneuvers compared to low- and medium-difficulty maneuvers (p<.001). Considering that higher RMSE reflects more error, and a higher expert rating reflect more accuracy, the results show that performance was generally worse in high-difficulty tasks, compared to low and medium difficulty levels, see trends in Figure 16.

Figure 16

Altitude Performance Across Difficulty Levels



Note. Error bars represent confidence intervals.

Emotional Intensity – Physiological Arousal

Difficulty levels had a significant effect on the frequency of SCR (Table 10). Pairwise comparisons revealed that participants experienced significantly more SCRs in high-difficulty compared to low-difficulty maneuvers (p=.002). See trends in Figure 17.

Table 10

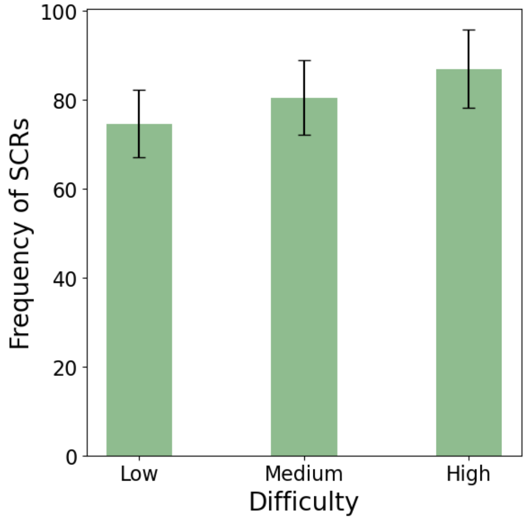
SCR Across Difficulty Levels

Emotion	Difficulty level	Mean	SD	MD	F	p	η^2
SCR	Low	74.61	18.56	L<H*	6.71	.003	.234
	Medium	80.48	20.41				
	High	86.96	21.40				

Note. SD=Standard Deviation, MD=Mean Difference, L=Low, M=Medium, H=High. *p=.017 after Bonferroni correction, **p<.001.

Figure 17

Skin Conductance Responses Across Difficulty Levels



Note. Error bars represent confidence intervals.

Emotional Intensity – Facial Expressions

Three RM-ANOVAs were conducted, one for each difficulty level (low-, medium-, and high-difficulty). In each RM-ANOVA, seven variables were compared, namely the frequency of single dominant emotions (i.e., anger, disgust, happy, disgust, fear, sad, and surprise); descriptive statistics in Table 11. The frequency of emotions had a statistically significant difference, see Figure 18. Participants showed more neutral, and anger facial expressions compared to happy, fear, and disgust for all three difficulty levels (low F=18.467, p<.001, η^2=.468, medium F=15.78, p<.001, η^2=.429, and high F=13.831, p<.001, η^2=.397). Dominant expressions of surprise were significantly more frequent than disgust in low-difficulty tasks, and dominant expressions of neutral were more frequent than sad in the three difficulty levels.

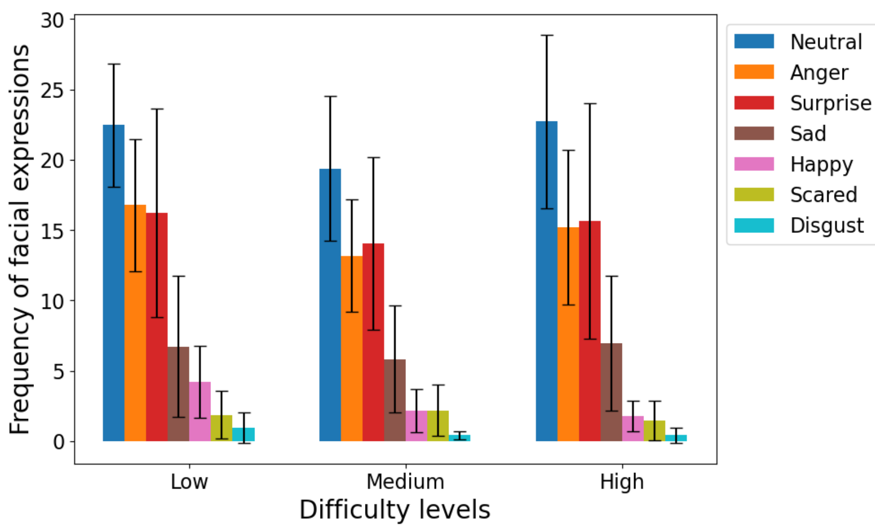
Table 11

Descriptive Statistics Frequency of Facial Expressions Across Difficulty Levels

	Low		Medium		Difficult	
	M	*SD*	*M*	*SD*	*M*	*SD*
Neutral	19.36	12.31	22.73	14.77	22.45	10.50
Anger	13.18	9.62	15.23	13.16	16.77	11.24
Surprise	14.04	14.73	15.64	20.08	16.23	17.78
Sad	5.81	9.14	6.95	11.45	6.73	11.98
Happy	2.14	3.67	1.77	2.56	4.23	6.13
Scared	2.18	4.10	1.45	3.33	1.86	4.05
Disgusted	0.41	0.73	0.41	1.30	0.95	2.63

Figure 18

Frequency of Facial Expressions across Difficulty Levels



Note. Pairwise comparisons: NE>SA in low*, medium* and high** difficulty levels. NE>HA, NE>SC, NE>DI in low**, medium** and high** difficulty levels. AN>HA in low*, medium*, and high* difficulty levels. AN>SC in low*, medium** and high** difficulty levels. AN>DI in low**, medium**, and high** difficulty levels. SU>DI in low* difficulty tasks. NE=Neutral, AN=Angry, SU=Surprise, SA=Sad, SC=Scared, DI=Disgust. *p<.007 after Bonferroni correction, **p<.001. Error bars represent confidence intervals.

As a follow up, RM-ANOVAs were conducted comparing single discrete emotions across difficulty levels. Results did not show a significant effect of difficulty level on expression for any of the seven emotions.

Emotional Variability – Fluctuation of Facial Expressions

Difficulty levels had a significant effect on emotional variability, descriptive statistics in Table 12. Pairwise comparisons revealed that emotional variability was marginally significant when comparing medium and high-difficulty tasks, showing that participants had a higher emotional variability in high-difficulty tasks compared to medium-difficulty tasks (p=.041). See trends in Figure 19.

Table 12

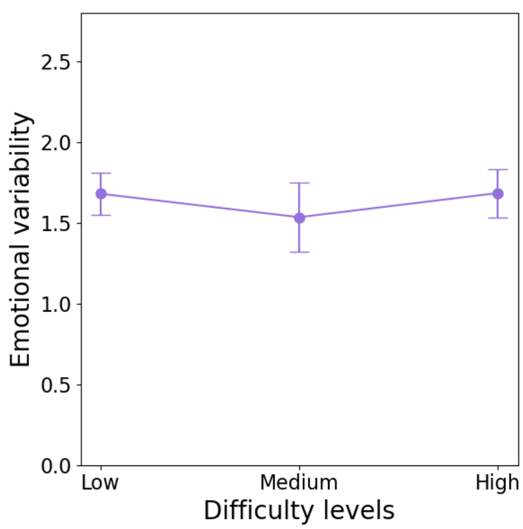
Emotional Variability Across Difficulty Levels

Performance measure	Difficulty level	Mean	SD	MD	F	p	η^2
Emotional variability	Low	1.68	.31		3.37	.044	.138
	Medium	1.53	.51	M<H*			
	High	1.68	.36				

Note. SD=Standard Deviation, MD=Mean Difference, L=Low, M=Medium, H=High. *p=.017 after Bonferroni correction, **p<.001.

Figure 19

Emotional Entropy Across Difficulty Levels



Note. Error bars represent confidence intervals.

Discussion

Differences Across Training Phases

Our results showed that performance changed across training phases. Aligned with our hypothesis, altitude control improved over training phases such that by the end of the experiment, participants were more accurate, requiring less amplitude corrections, and were granted a better score by the expert. However, heading error and expert rating did not have a statistically significant difference across phases. A potential explanation is that the yoke movements to control heading are similar to performing car turns using a steering wheel. Thus, participants were likely more familiar with the movement to control heading, causing less error and less differences across training phases. In contrast, since participants were mostly brand-new beginners, they might had focused on altitude, as an unknown metric, and had less intuitive control. This is a known pattern among novices who tend to concentrate on one aspect of the task when it is not familiar, compared to experts who have the ability to visualize problems from a global perspective, and have perfected technical methods and crosscheck techniques (Ericsson, 2006; Lajoie & Gube, 2018).

When analyzing changes on high intensity (single-variable emotions), we found significant changes in skin conductance responses and frequency of sadness. Our expectations regarding changes in physiological arousal were partially fulfilled. Aligned with our hypothesis, participants experienced more skin conductance responses in session A compared to session B. Based on this finding, we infer that participants were most engaged during session A, which aligns with experiencing more boredom (and less engagement) at the end of the task (Goetz & Hall, 2014; Hidalgo-Munoz et al., 2018; Koglbauer et al., 2011). However, in our hypothesis we did not clarify expectations regarding differences between introduction and session A, finding that participants experienced more skin conductance responses in session A. In line with our hypothesis, sadness was significantly higher in session B compared to introduction phase. Since facial expression of sadness can be mapped onto boredom (Ahn & Harley, 2020; Harley et al., 2015), participants might have experienced boredom at the end of the experiment due to the time investment, and loss of motivation as time passed (Hidalgo-Munoz et al., 2018; Rosa et al., 2021, 2022). A potential explanation of this pattern can be that feedback was removed as the practice phases begun. In our experimental design, researchers provided only constructive feedback during the training phase (Krahenbuhl et al., 1981). The positive feedback served as a learning scaffold for participants to fill the gap between current and potential knowledge (Lajoie, 2017). Therefore, participants were likely to be less aroused when receiving feedback (during introduction) compared to when they performed the task independently for the first time (session A). As participants habituated to the task and its expectations (during session B), there was a decrease in physiological arousal and emotional variability (Hidalgo-Munoz et al., 2018; Krahenbuhl et al., 1981).

When analysing differences in intense emotions, inferred from facial expressions, we found that emotions had a binary separation between high vs. low frequency. Namely, neutral and anger were significantly more frequent across the three training phases compared to sadness, happiness, fear, and disgust, in that order. As expected, neutral served as an emotional baseline, thus, it is likely that trainees were not emotionally distracted during the task and were able to engage their cognitive efforts to perform the flying maneuvers (Harley et al., 2012), which at the same time can explain the improvement in altitude performance by the end of the task.

However, explaining the frequency of anger was less straight forward. Anger, paired with frustration (Harley et al., 2015; Pekrun, 2019), is a negative-deactivating emotion that occurs when the learner perceives the task demands as unreasonable or controlled by external factors (Pekrun & Perry, 2014). In a simple explanation, participants might had perceived the task as unnecessarily complicated, however they were determined (aroused) to finish the task. However, the current analysis lacks information as of the locus of control of participants. Therefore, we propose a second explanation. Facial expression of anger, specially lowering browns and tightening the eye lids (Farnsworth, 2022), has been previously interpreted as deep-focus in a flying task, especially when looking at a screen with a "frowning face" (T. Li & Lajoie, 2021). Due to the task characteristics, since participants are facing a screen and it is a low-stakes task for learning how to use a simulator, we argue that the second interpretation might be more pertinent.

The interpretations of our results are linked to our measurements. FaceReader 6.0 is built as an objective measure detecting facial action units that are then merged and results are labelled according to basic emotions theory and the circumplex model of affect (Ekman, 1992, den Uyl et al., 2023; Russell, 2003). FaceReader's manufacturer, Noldus, has recently changed its approach, encouraging users to interpret facial expressions according to their needs and theoretical standing (den Uyl et al., 2023). Therefore, a limitation of using FaceReader version 6.0 is that it does not have personalization features to label groups of action units, as in our case anger or deep focus, leading us to interpret a frowning face according to our context, specifically ab-initio pilot trainees interacting with a flight simulator (den Uyl et al., 2023; Loijens et al., 2015). We recognize this limitation; however, our results invite future research to use more flexible facial-detection software, such as FaceReader 9.0 or OpenFace, to label emotions according to each study theoretical framework. Results did not show a statistically significant difference in emotional variability across training phases, but a trend showed that emotional variability was higher during session A than during introduction and session B training phases.

Regarding emotional intensity, we recognize a few limitations. The non-normality of the facial expression of the non-frequent emotions (i.e., surprise, fear, disgust, happiness, and surprise) may impact the generalizability of our findings, and emotional variability did not show statistically significant differences and a low effect size (Cohen, 1988), and the interpretations of our results are linked to our measurements. FaceReader 6.0 is built as an objective measure detecting facial action units that are merged and labelled according to basic emotions theory (Ekman, 1992, den Uyl et al., 2023). FaceReader's manufacturer, Noldus, has recently changed its approach, encouraging users to interpret facial expressions according to their needs and theoretical standing (den Uyl et al., 2023). Therefore, a limitation of using FaceReader version 6.0 is that it does not have personalization features to label groups of action units, as in our case anger or deep focus, leading us to interpret a frowning face according to our context, specifically ab-initio pilot trainees interacting with a flight simulator (den Uyl et al., 2023; Loijens et al., 2015). We recognize this limitation; however, our results invite future research to use more flexible facial-detection software, such as FaceReader 9.0 or OpenFace, to label emotions according to each study theoretical framework.

We encourage future directions according to our results. Future studies and pilots training curricular design account for increases in trainees' boredom (D'Mello & Graesser, 2012; Goetz & Hall, 2014). For instance, instructional methods can reduce long working hours, which is known to associate with human error in aviation (Code of Federal Regulations, 2023; Wingelaar-Jagt et al., 2021). Moreover, future studies could explore the correlation of neutral expressions as a positive correlation with flying performance. Previous studies have explored undergraduate students' emotions when interacting with intelligent tutoring systems, similarly finding that neutral is one of the most frequent emotions that learners experience (D'Mello & Graesser, 2012; Harley et al., 2013). Contrary to our results, these studies found that confusion (paired with disgust) and happiness were identified as the most frequent emotions (accordingly D'Mello & Graesser, 2012; Harley et al., 2013). Consequently, more studies are needed to explore if the frequency of emotions is aligned to the profession (i.e., pilots) and/or the task (i.e., simulated flying).

In summary, the introduction phase was characterized as having poorer altitude performance and lower frequency of emotions overall, with participants experiencing less sadness and less skin conductance responses. During session A, performance improved being significantly better than introduction but not different from session B; additionally, participants experienced more skin conductance responses, and we observed (descriptively) more emotional variability. Lastly, during session B performance was more accurate and participants had a higher frequency of neutral and sad expressions across the three phases.

Differences Across Difficulty Levels

Our results confirmed that flying performance and emotions varied across difficulty levels. The trend in performance was clear, showing that flying performance was less precise in high-difficulty tasks compared to low and medium difficulty levels.

Aligned with our expectations, participants experienced more emotional intensity, inferred from more skin conductance responses, during high-difficulty maneuvers than low-difficulty maneuvers. These findings align with previous research showing that pilot trainees' peaks of physiological arousal increase with task difficulty (Gaetan et al., 2015; Skibniewski et al., 2015). Moreover, more skin conductance responses indicates that participants were more emotionally and cognitively engaged when attempting to solve the task (Braithwaite et al., 2015; Harley et al., 2019b).

When analysing the differences of emotional intensity across difficulty levels, neutral remains the most frequently expressed dominant emotion, being significantly more frequent than sadness, happiness, fear, and disgust, confirming that neutral serves as a baseline state and shows that participants are generally stable during the flying task (Harley et al., 2012). Contrary to the analysis across phases, the results add a third layer of moderately frequent emotions. Particularly, anger is not statistically different from sadness in any level, and surprise is significantly more frequent than disgust in low-difficulty maneuvers. These findings may be due to the split in the statistical analysis. For analysing differences across difficulty levels, we used trials 8-22 for having an even number of tasks per level, whereas changes across training phases included all trials. Particularly, the introduction had more low-difficulty cases (i.e., four out of seven), and participants received feedback on their flying performance. Thus, in this second analysis, by focusing on session A and B, it is likely that participants were less concentrated (expressing less anger) and more surprised by the novelty of the task as they had to perform without researchers' feedback for the first time (Landman et al., 2017; T. Li & Lajoie, 2021).

Moreover, research in flight training has had a deeper interest in understanding negative-activating affect like stress and anxiety (Allsop & Gray, 2016; Hart, 2006). However, our results suggest that anxiety, paired with facial expression of fear, was not frequently expressed by participants. These results suggest that the prevalence and function of a wider range of emotions should be explored in this context.

Emotional variability was higher during high-difficulty tasks compared to medium-difficulty levels, similar to results found with medical students (S. Li, et al., 2021a). Difficult tasks required more changes between aircraft metrics, and these constant changes might imply constant fluctuations in object focus, triggering different emotions and causing less stability (Pekrun, 2019). However, results show that emotional variability was (descriptively) higher during low than medium-difficulty tasks. Our results show a U-shaped pattern between emotional variability and difficulty level, in which emotional variability was higher in low and difficult tasks, compared to medium difficulty; yet performance was increasingly worse as difficulty level increased. Emotional variability might be following the Yerkes-Dodson law (1908), in which an optimal level of emotional fluctuations can improve performance, but trespassing that threshold (as in the high-difficulty tasks) might imply an impairment in flying accuracy. This threshold might be inferred by physiological arousal which was higher in high than low-difficulty tasks.

It is recognized that our analyses had limitations that could guide future research. Participants were not enrolled as pilot students, but rather represent ab-initio pilot trainees (i.e., brand-new trainees) with limited experience using flight simulators or aircraft. Although we confirmed that participants' previous experience using flight simulators did not influence performance and emotional reactions, educational background played a role when comparing difficulty levels: participants with master's degree had higher altitude RMSE during medium-difficulty tasks, expressed less anger during low-difficulty tasks, and were more neutral during high-difficulty tasks compared to participants with a bachelor's degree. Aligned with the CVT, we believe that participants who had invested more in their professional formation, having a higher educational degree, could have had a lower perceived control and value over the tasks, since acting as pilot trainees is likely farther from their career path. Consequently, we suggest future studies to replicate or extend the test methodology and include student pilots recently enrolled in aviation school and more advanced pilots. Moreover, it is suggested that the analysis is reproduced with a larger sample size to increase statistical power. We recommend that future studies record and control for instructors' feedback to understand how advice aligns with students' expectations and emotional responses (Naismith & Lajoie, 2018; Shute, 2008).

Moreover, when analysing changes of single facial expressions across difficulty levels there were no statistically significant differences for any of the seven emotions. Only expression of happiness had a marginal statistical difference, being more frequent during high-difficulty than low and medium-difficulty tasks. In that sense, we recognize a common critique of using behavioral measures to evaluate emotion: behavioral responses should be triangulated with people's subjective experience, such as questionnaires or interviews (Harley, 2016). Nevertheless, this was a calculated risk to understand the feasibility of using a non-invasive and non-distracting measure of emotions, which is particularly relevant in the context of aviation training where pilots require full concentration on the task, rather than responding to questionnaires concurrently during flying tasks.

For future studies, we suggest exploring emotional variability of pilot trainees in a larger sample and adding (non-interrupting) self-report measures to understand trainees' subjective experience to identify the interaction between emotional variability and flying performance. In the current study we explored behavioral and physiological expression of emotions during the same time period; however, we suggest that future research explores the co-occurrence of behavioral and physiological expression of emotions, such as checking time points in which the significant facial expressions co-occur with skin conductance responses in key moments of the task (i.e., sudden decrease of altitude) (Ruiz Segura, 2020). Such combinations could be automatically detected to create interventions for trainees to up or down-regulate their emotions, resulting in optimal performance (Gross, 2015). Additionally, future research can explore more emotional patterns, such as sequences and co-occurrences of emotions (Zheng et al., 2023).

In summary, in low and medium difficulty levels, performance was significantly better than in difficult tasks, and participants had more neutral than sad expressions. In low-difficulty tasks, participants expressed more surprise than disgust, and had less skin conductance responses than in difficult tasks. Lastly, high-difficulty tasks had worse performance than low and medium-difficulty levels, participants expressed overall more emotions, had more skin conductance responses, and had more emotional variability.

Conclusion

In this study we argue that a means to improve pilot trainees experience and performance is to account for their emotional experiences when flying (Martins, 2016). This study shows that pilot trainees experience dynamic changes among multiple emotions. Our results show that emotion dynamics, such as intensity, frequency, and variability. and flying performance change according to training phases and difficulty levels (Bailen et al., 2019; Zheng et al., 2023a). As a mean to improve the quality of learning and performance.

This manuscript serves as a baseline to understand emotional experiences of brand-new pilot training to inform the creation of interventions to improve flight performance. To reduce human error, pilot trainees can learn emotion regulation techniques as part of their curriculum. Particularly, future training can use simulated flying tasks to assist trainees to familiarize themselves with the emotional reactions they have as training advances or encountering difficult tasks. More awareness about their emotions might help pilot trainees recognize key moments of disengagement or high physiological arousal, thus allowing trainees to self-regulate and modify their behaviour or strategy or permitting instructors to intervene in a timely fashion, and reduce likelihood of errors when flying an airplane (Sieberichs & Kluge, 2018).

References

Ahn, B., & Harley, J. M. (2020). Facial expressions when learning with a queer history app: Application of the control value theory of achievement emotions. *British Journal of Educational Technology*, *51*(5), 1563–1576. https://doi.org/10.1111/bjet.12989

Allsop, J., & Gray, R. (2014). Flying under pressure: Effects of anxiety on attention and gaze behavior in aviation. *Journal of Applied Research in Memory and Cognition*, *3*(2), 63–71. https://doi.org/10.1016/j.jarmac.2014.04.010

Artino, A. R., Holmboe, E. S., & Durning, S. J. (2012). Control-value theory: Using achievement emotions to improve understanding of motivation, learning, and performance in medical education: AMEE Guide No. 64. *Medical Teacher*, *34*(3), e148–e160. https://doi.org/10.3109/0142159X.2012.651515

Azher, S., Cervantes, A., Marchionni, C., Grewal, K., Marchand, H., & Harley, J. M. (2023). Virtual simulation in nursing education: Headset virtual reality and screen-based virtual simulation offer a comparable experience. *Clinical Simulation in Nursing*, *79*, 61–74. https://doi.org/10.1016/j.ecns.2023.02.009

Bailen, N. H., Green, L. M., & Thompson, R. J. (2019). Understanding emotion in adolescents: A review of emotional frequency, intensity, instability, and clarity. *Emotion Review*, *11*(1), 63–73. https://doi.org/10.1177/1754073918768878

Baumann, M. R., Gohm, C. L., & Bonner, B. L. (2011). Phased training for high-reliability occupations: Live-fire exercises for civilian firefighters. *Human Factors*, *53*(5), 548–557. https://doi.org/10.1177/0018720811418224

Berner, E. S., & Graber, M. L. (2008). Overconfidence as a cause of diagnostic error in medicine. *The American Journal of Medicine*, *121*(5 Suppl), S2-23. https://doi.org/10.1016/j.amjmed.2008.01.001

Boucsein, W., Fowles, D. C., Grimnes, S., Ben-Shakhar, G., Walton. T. R., Dawson, M. E., & Filion, D. L. (2012). Publication recommendations for electrodermal measurements. *Psychophysiology*, *49*(8), 1017–1034. https://doi.org/10.1111/j.1469-8986.2012.01384.x

Braithwaite, J. J., Watson, D. G., Jones, R., & Rowe, M. (2015). *A guide for analysing electrodermal activity (EDA) and skin conductance responses (SCRs) for psychological experiments*. SAAL, Behavioural Brain Sciences Centre, University of Birminghan. https://www.birmingham.ac.uk/Documents/college-les/psych/saal/guide-electrodermal-activity.pdf

Code of Federal Regulations. (2023, April 19). *14 CFR 91.1059 -- Flight time limitations and rest requirements: One or two pilot crews.* Electronic code of federal regulations. https://www.ecfr.gov/current/title-14/chapter-I/subchapter-F/part-91/subpart-K/subject-group-ECFRc17623c0e0be17e/section-91.1059

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Routledge. https://doi.org/10.4324/9780203771587

Cootes, T. F., & Taylor, C. J. (2001). Statistical models of appearance for medical image analysis and computer vision. *Medical Imaging 2001: Image Processing*, *4322*, 236–248. https://doi.org/10.1117/12.431093

D'Mello, S., & Graesser, A. (2012). Dynamics of affective states during complex learning. *Learning and Instruction*, *22*(2), 145–157. https://doi.org/10.1016/j.learninstruc.2011.10.001

D'Mello, S., Lehman, B., & Person, N. (2010). Monitoring affect states during effortful problem solving activities. *International Journal of Artificial Intelligence in Education*, *20*(4), 361–389. https://doi.org/10.3233/JAI-2010-012

Duffy, M. C., Lajoie, S., & Lachapelle, K. (2016). Measuring emotions in medical education: Methodological and technological advances within authentic medical learning environments. In S. Bridges, L. K. Chan, & C. E. Hmelo-Silver (Eds.), *Educational Technologies in Medical and Health Sciences Education* (pp. 181–213). Springer. https://doi.org/10.1007/978-3-319-08275-2_10

Ekman, P. (1992). An argument for basic emotions. *Cognition and Emotion*, *6*(3–4), 169–200. https://doi.org/10.1080/02699939208411068

Ekman, P., & Rosenberg, E. L. (Eds.). (2005). *What the face reveals: Basic and applied studies of spontaneous expression using the facial action coding system (FACS)* (2nd ed.). Oxford University Press. 10.1093/acprof:oso/9780195179644.001.0001

Endsley, M. R. (2000). Theoretical underpinnings of situation awareness: A critical review. In M. R. Endsley & D. J. Garland (Eds.), *Situation Awareness Analysis and Measurement* (pp. 3-28). Lawrence Erlbaum Associates.

Ericsson, K. A. (2006). The influence of experience and deliberate practice on the development of superior expert performance. In K. A. Ericsson, N. Charness, P. J. Feltovich, & R. R. Hoffman (Eds.), *The Cambridge handbook of expertise and expert performance* (pp. 683–704). Cambridge University Press. https://doi.org/10.1017/CBO9780511816796.038

European Commission. (2013, October 4). *Pilot and crew fatigue—Frequently asked questions*. Directorate-General for Communication, European Commission. https://ec.europa.eu/commission/presscorner/detail/en/MEMO_13_854

FAA Department of Transportation. (2003, September 17). *14 CFR 91.1059 -- Flight time limitations and rest requirements: One or two pilot crews.* Electronic Code of Federal Regulations. https://www.ecfr.gov/current/title-14/chapter-I/subchapter-F/part-91/subpart-K/subject-group-ECFRc17623c0e0be17e/section-91.1059

Farnsworth, B. (2022, October 18). Facial Action Coding System (FACS)—A Visual Guidebook—*iMotions*. https://imotions.com/blog/learning/research-fundamentals/facial-action-coding-system/

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*(2), 175–191. https://doi.org/10.3758/bf03193146

Gaba, D. M. (2004). The future vision of simulation in health care. *Quality & Safety in Health Care*, *13 Suppl 1*(Suppl 1), i2-10. https://doi.org/10.1136/qhc.13.suppl_1.i2

Gaetan, S., Dousset, E., Marqueste, T., Bringoux, L., Bourdin, C., Vercher, J.-L., & Besson, P. (2015). Cognitive workload and psychophysiological parameters during multitask activity in helicopter pilots. *Aerospace Medicine and Human Performance*, *86*(12), 1052–1057. https://doi.org/10.3357/AMHP.4228.2015

Goetz, T., & Hall, N. C. (2014). Academic boredom. In R. Pekrun & L. Linnenbrink-Garcia (Eds.), *International handbook of emotions in education* (pp. 311–330). Routledge.

Gray, R., Gaska, J., & Winterbottom, M. (2016). Relationship between sustained, orientated, divided, and selective attention and simulated aviation performance: Training & pressure effects. *Journal of Applied Research in Memory and Cognition*, *5*(1), 34–42. https://doi.org/10.1016/j.jarmac.2015.11.005

Gross, J. J. (2015). Emotion regulation: Current status and future prospects. *Psychological Inquiry*, *26*(1), 1–26. https://doi.org/10.1080/1047840X.2014.940781

Hamman, W. R. (2004). Will simulation fly in medicine as it has in aviation? *Quality & Safety in Health Care*, *13*(5), 397–399. https://doi.org/10.1136/qhc.13.5.397

Han, S.-Y., Kwak, N.-S., Oh, T., & Lee, S.-W. (2020). Classification of pilots' mental states using a multimodal deep learning network. *Biocybernetics and Biomedical Engineering*, *40*(1), 324–336. https://doi.org/10.1016/j.bbe.2019.12.002

Hardy, M., Wiebe, E. N., Grafsgaard, J. F., Boyer, K. E., & Lester, J. C. (2013). Physiological responses to events during training: Use of skin conductance to inform future adaptive learning systems. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *57*(1), 2101–2105. https://doi.org/10.1177/1541931213571468

Harley, J. M. (2016). Measuring emotions: A survey of cutting edge methodologies used in computer-based learning environment research. In S. Y. Tettegah & M. Gartmeier (Eds.), *Emotions, technology, design, and learning* (pp. 89–114). Academic Press. https://doi.org/10.1016/B978-0-12-801856-9.00005-0

Harley, J. M., & Azevedo, R. (2014). Toward a feature-driven understanding of students' emotions during interactions with agent-based learning environments: A selective review. *International Journal of Gaming and Computer-Mediated Simulations*, *6*(3), 17–34. https://doi.org/10.4018/ijgcms.2014070102

Harley, J. M., Bilgic, E., Lau, C. H. H., Gorgy, A., Marchand, H., Lajoie, S. P., Lavoie-Tremblay, M., & Fried, G. M. (2023). Nursing students reported more positive emotions about training during Coronavirus disease 2019 (COVID-19) after using a virtual simulation paired with an in-person simulation. *Clinical Simulation in Nursing*, *81*, 101420. https://doi.org/10.1016/j.ecns.2023.04.006

Harley, J. M., Bouchet, F., & Azevedo, R. (2012). Measuring learners' co-occurring emotional responses during their interaction with a pedagogical agent in MetaTutor. In S. A. Cerri, W. J. Clancey, G. Papadourakis, & K. Panourgia (Eds.), *Intelligent tutoring systems* (pp. 40–45). Springer. https://doi.org/10.1007/978-3-642-30950-2_5

Harley, J. M., Bouchet, F., & Azevedo, R. (2013). Aligning and comparing data on emotions experienced during learning with MetaTutor. In H. C. Lane, K. Yacef, J. Mostow, & P. Pavlik (Eds.), *Artificial intelligence in education* (pp. 61–70). Springer. https://doi.org/10.1007/978-3-642-39112-5_7

Harley, J. M., Bouchet, F., Hussain, M. S., Azevedo, R., & Calvo, R. (2015). A multi-componential analysis of emotions during complex learning with an intelligent multi-agent system. *Computers in Human Behavior*, *48*, 615–625. https://doi.org/10.1016/j.chb.2015.02.013

Harley, J. M., Jarrell, A., & Lajoie, S. P. (2019b). Emotion regulation tendencies, achievement emotions, and physiological arousal in a medical diagnostic reasoning simulation. *Instructional Science*, *47*(2), 151–180. https://doi.org/10.1007/s11251-018-09480-z

Harley, J. M., Pekrun, R., Taxer, J. L., & Gross, J. J. (2019a). Emotion regulation in achievement situations: An integrated model. *Educational Psychologist*, *54*(2), 106–126. https://doi.org/10.1080/00461520.2019.1587297

Hart, S. G., & Bortolussi, M. R. (1984). Pilot errors as a source of workload. *Special Issue: Aviation Psychology*, *26*(5), 545–556.

Herrera-Aliaga, E., & Estrada, L. D. (2022). Trends and innovations of simulation for twenty first century medical education. *Frontiers in Public Health*, *10*, 619769. https://doi.org/10.3389/fpubh.2022.619769

Hidalgo-Munoz, A. R., Mouratille, D., Matton, N., Causse, M., Rouillard, Y., & El-Yagoubi, R. (2018). Cardiovascular correlates of emotional state, cognitive workload and time-on-task effect during a realistic flight simulation. *International Journal of Psychophysiology*, *128*, 62–69. https://doi.org/10.1016/j.ijpsycho.2018.04.002

Hunter, J. D. (2007). Matplotlib: A 2D graphics environment. *Computing in Science & Engineering*, *9*(3), 90–95. https://doi.org/10.1109/MCSE.2007.55

Jack, R. E., Garrod, O. G. B., & Schyns, P. G. (2014). Dynamic facial expressions of emotion transmit an evolving hierarchy of signals over time. *Current Biology: CB*, *24*(2), 187–192. https://doi.org/10.1016/j.cub.2013.11.064

Jennings, S., Law, A., Bourgon, A., & Grenier, H. (2024, January). *Novel methodology for comparing student pilot flight performance with instructor pilot expert ratings*. AIAA SCITECH 2024 Forum, Orlando, Florida.

Karaca, Y., & Moonis, M. (2022). Shannon entropy-based complexity quantification of nonlinear stochastic process: Diagnostic and predictive spatiotemporal uncertainty of multiple sclerosis subgroups. In Y. Karaca, D. Baleanu, Y.-D. Zhang, O. Gervasi, & M. Moonis (Eds.), *Multi-Chaos, Fractal and Multi-Fractional Artificial Intelligence of Different Complex Systems* (pp. 231–245). Academic Press. https://doi.org/10.1016/B978-0-323-90032-4.00018-3

Kinney, L., & O'Hare, D. (2020). Responding to an unexpected in-flight event: Physiological arousal, information processing, and performance. *Human Factors*, *62*(5), 737–750. https://doi.org/10.1177/0018720819854830

Koglbauer, I., Kallus, K., Braunstingl, R., & Boucsein, W. (2011). Recovery training in simulator improves performance and psychophysiological state of pilots during simulated and real visual flight rules flight. *The International Journal of Aviation Psychology*, *21*(4), 307–324. https://doi.org/10.1080/10508414.2011.606741

Krahenbuhl, G. S., Darst, P. W., Marett, J. R., Reuther, L. C., Constable, S. H., Swinford, M. E., & Reid, G. B. (1981). Instructor pilot teaching behavior and student pilot stress in flight training. *Aviation, Space, and Environmental Medicine*, *52*(10), 594–597.

Krone, T., Albers, C. J., Kuppens, P., & Timmerman, M. E. (2018). A multivariate statistical model for emotion dynamics. *Emotion*, *18*(5), 739–754. https://doi.org/10.1037/emo0000384

Lajoie, S. P. (2017). Learning science applications for research in medicine. In L. Lin & J. M. Spector (Eds.), *The sciences of learning and instructional design* (pp. 108–118). Routledge. https://doi.org/10.4324/9781315684444-8

Lajoie, S. P. (2021). Multimedia learning with simulations. In L. Fiorella & R. E. Mayer (Eds.), *The Cambridge handbook of multimedia learning* (3rd ed., pp. 461–471). Cambridge University Press. https://doi.org/10.1017/9781108894333.048

Lajoie, S. P., & Gube, M. (2018). Adaptive expertise in medical education: Accelerating learning trajectories by fostering self-regulated learning. *Medical Teacher*, *40*(8), 809–812. https://doi.org/10.1080/0142159X.2018.1485886

Lajoie, S. P., Li, S., & Zheng, J. (2023). The functional roles of metacognitive judgement and emotion in predicting clinical reasoning performance with a computer simulated environment. *Interactive Learning Environments*, *31*(6), 3464–3475. https://doi.org/10.1080/10494820.2021.1931347

Lajoie, S. P., Naismith, L., Poitras, E., Hong, Y.-J., Cruz-Panesso, I., Ranellucci, J., Mamane, S., & Wiseman, J. (2013). Technology-rich tools to support self-regulated learning and performance in medicine. In R. Azevedo & V. Aleven (Eds.), *International handbook of metacognition and learning technologies* (pp. 229–242). Springer. https://doi.org/10.1007/978-1-4419-5546-3_16

Lajoie, S. P., Zheng, J., Li, S., Jarrell, A., & Gube, M. (2021). Examining the interplay of affect and self-regulation in the context of clinical reasoning. *Learning and Instruction*, *72*, 101219. https://doi.org/10.1016/j.learninstruc.2019.101219

Landaman, A., Groen, E. L., van Paassen, M. M., Bronkhorst, A. W., Mulder, M., & Mulder, M. (2017). Dealing with unexpected events on the flight deck: A conceptual model of startle and surprise. *Human Factors*, *59*(8), 1161–1172. https://doi.org/10.1177/0018720817723428

Landman, A., van Middelaar, S. H., Groen, E. L., van Paassen, M. M., Bronkhorst, A. W., & Mulder, M. (2020). The effectiveness of a mnemonic-type startle and surprise management procedure for pilots. *The International Journal of Aerospace Psychology*, *30*(3–4), 104–118. https://doi.org/10.1080/24721840.2020.1763798

Lazarus, R. S., PhD, & Folkman, S., PhD. (1984). *Stress, appraisal, and coping*. Springer. 10.1007/978-1-4419-1005-9_215

Lesne, A. (2014). Shannon entropy: A rigorous notion at the crossroads between probability, information theory, dynamical systems and statistical physics. *Mathematical Structures in Computer Science*, *24*(3), e240311. https://doi.org/10.1017/S0960129512000783

Lewinski, P., den Uyl, T. M., & Butler, C. (2014). Automated facial coding: Validation of basic emotions and FACS AUs in FaceReader. *Journal of Neuroscience, Psychology, and Economics*, *7*(4), 227–236. https://doi.org/10.1037/npe0000028

Li, S., Zheng, J., & Lajoie, S. P. (2021). The frequency of emotions and emotion variability in self-regulated learning: What matters to task performance? *Frontline Learning Research*, *9*(4), https://doi.org/10.14786/flr.v9i4.901

Li, S., Zheng, J., Lajoie, S. P., & Wiseman, J. (2021). Examining the relationship between emotion variability, self-regulated learning, and task performance in an intelligent tutoring system. *Educational Technology Research and Development*, *69*(2), 673–692. https://doi.org/10.1007/s11423-021-09980-9

Li, T., & Lajoie, S. (2021). Predicting aviation training performance with multimodal affective inferences. *International Journal of Training and Development*, *25*(3), 301–315. https://doi.org/10.1111/ijtd.12232

Loijens, L., & Krips, O. (2021). *FaceReader methodology note*. Noldus Information Technology.

Loijens, L., Krips, O., van Kuilenburg, H., den Uyl, M., & Ivan, P. (2015). *FaceReader version 6.1 Reference manual*. Noldus Information Technology.

Makowski, D., Pham, T., Lau, Z. J., Brammer, J. C., Lespinasse, F., Pham, H., Schölzel, C., & Chen, S. H. A. (2021). NeuroKit2: A Python toolbox for neurophysiological signal processing. *Behavior Research Methods*, *53*(4), 1689–1696. https://doi.org/10.3758/s13428-020-01516-y

Marques, E., Carim, G., Campbell, C., & Lohmann, G. (2023). Ab initio flight training: A Systematic literature review. *The International Journal of Aerospace Psychology*, *33*(2), 99–119. https://doi.org/10.1080/24721840.2022.2162405

Martins, A. P. G. (2016). A review of important cognitive concepts in aviation. *Aviation*, *20*(2), 65–84. https://doi.org/10.3846/16487788.2016.1196559

Naismith, L. M., & Lajoie, S. P. (2018). Motivation and emotion predict medical students' attention to computer-based feedback. *Advances in Health Sciences Education*, *23*(3), 465–485. https://doi.org/10.1007/s10459-017-9806-x

Nomura, O., Wiseman, J., Sunohara, M., Akatsu, H., & Lajoie, S. P. (2021). Japanese medical learners' achievement emotions: Accounting for culture in translating Western medical educational theories and instruments into an asian context. *Advances in Health Sciences Education*, *26*(4), 1255–1276. https://doi.org/10.1007/s10459-021-10048-9

Pekrun, R. (2019). Achievement emotions: A control-value theory perspective. In R. Patulny, A. Bellocchi, R. Olson, S. Khorana, J. McKenzie, & M. Peterie (Eds.), *Emotions in late modernity* (pp. 142-157). Routledge.

Pekrun, R., & Perry, R. P. (2014). Control-value theory of achievement emotions. In R. Pekrun & L. Linnenbrink-Garcia (Eds.), *International handbook of emotions in education* (pp. 120–141). Routledge.

Rosa, E., Gronkvist, M., Kolegard, R., Dahlstrom, N., Knez, I., Ljung, R., & Willander, J. (2021). Fatigue, emotion, and cognitive performance in simulated long-duration, single-piloted flight missions. *Aerospace Medicine and Human Performance*, *92*(9), 710–719. https://doi.org/10.3357/AMHP.5798.2021

Rosa, E., Lyskov, E., Gronkvist, M., Kolegard, R., Dahlstrom, N., Knez, I., Ljung, R., & Willander, J. (2022). Cognitive performance, fatigue, emotional, and physiological strains in simulated long-duration flight missions. *Military Psychology*, *34*(2), 224–236. https://doi.org/10.1080/08995605.2021.1989236

Ruiz Segura, A. (2020). *Examining the interaction between clinical reasoning and emotion on medical students' diagnostic efficiency* [Masters thesis, McGill University]. https://escholarship.mcgill.ca/concern/theses/5h73q1579

Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology*, *39*(6), 1161–1178. https://doi.org/10.1037/h0077714

Russell, J. A. (2003). Core affect and the psychological construction of emotion. *Psychological Review*, *110*(1), 145–172. https://doi.org/10.1037/0033-295X.110.1.145

Sassenus, S., Van den Bossche, P., & Poels, K. (2022). When stress becomes shared: Exploring the emergence of team stress. *Cognition, Technology & Work*, *24*, 537–556. https://doi.org/10.1007/s10111-022-00698-z

Shannon, C. E. (1948). A mathematical theory of communication. *The Bell System Technical Journal*, *27*(3), 379–423. https://doi.org/10.1002/j.1538-7305.1948.tb01338.x

Shute, V. J. (2008). Focus on formative feedback. *Review of Educational Research*, *78*(1), 153–189. https://doi.org/10.3102/0034654307313795

Sieberichs, S., & Kluge, A. (2018). Effects of in-flight countermeasures to mitigate fatigue risks in aviation. *Aviation Psychology and Applied Human Factors*, *8*(2), 86–92. https://doi.org/10.1027/2192-0923/a000143

Skibniewski, F. W., Dziuda, L., Baran, P. M., Krej, M. K., Guzowski, S., Piotrowski, M. A., & Truszczynski, O. E. (2015). Preliminary results of the LF/HF ratio as an indicator for estimating difficulty level of flight tasks. *Aerospace Medicine and Human Performance*, *86*(6), 518–523. https://doi.org/10.3357/AMHP.4087.2015

Su, C., Li, X., Yang, L., & Zeng, Y. (2022). *Mechanism, measurement, and quantification of stress in decision process: A model based systematic-review protocol* (arXiv:2203.10397). arXiv. https://doi.org/10.48550/arXiv.2203.10397

Thompson, R. J., Mata, J., Jaeggi, S. M., Buschkuehl, M., Jonides, J., & Gotlib, I. H. (2012). The everyday emotional experience of adults with major depressive disorder: Examining emotional instability, inertia, and reactivity. *Journal of Abnormal Psychology*, *121*(4), 819–829. https://doi.org/10.1037/a0027978

Tichon, J. G., Wallis, G., Riek, S., & Mavin, T. (2014). Physiological measurement of anxiety to evaluate performance in simulation training. *Cognition, Technology & Work*, *16*(2), 203–210.

Transport Canada. (2019, November 4). *Standard 421—Flight Crew Permits, Licences and Ratings—Canadian Aviation Regulations (CARs)*. AARBH 15472583; AARBH. https://tc.canada.ca/en/corporate-services/acts-regulations/list-regulations/canadian-aviation-regulations-sor-96-433/standards/standard-421-flight-crew-permits-licences-ratings-canadian-aviation-regulations-cars

Trull, T. J., Solhan, M. B., Tragesser, S. L., Jahng, S., Wood, P. K., Piasecki, T. M., & Watson, D. (2008). Affective instability: Measuring a core feature of borderline personality disorder with ecological momentary assessment. *Journal of Abnormal Psychology*, *117*(3), 647–661. https://doi.org/10.1037/a0012532

Vine, S. J., Uiga, L., Lavric, A., Moore, L. J., Tsaneva-Atanasova, K., & Wilson, M. R. (2015). Individual reactions to stress predict performance during a critical aviation incident. *Anxiety, Stress & Coping: An International Journal*, *28*(4), 467–477.

Viola, P., & Jones, M. J. (2004). Robust real-time face detection. *International Journal of Computer Vision*, *57*(2), 137–154. https://doi.org/10.1023/B:VISI.0000013087.49260.fb

Wingelaar-Jagt, Y. Q., Wingelaar, T. T., Riedel, W. J., & Ramaekers, J. G. (2021). Fatigue in aviation: Safety risks, preventive strategies and pharmacological interventions. *Frontiers in Physiology*, *12*, 712628. https://doi.org/10.3389/fphys.2021.712628

Yerkes, R. M., & Dodson, J. D. (1908). The relation of strength of stimulus to rapidity of habit formation. *Journal of Comparative Neurology & Psychology*, *18*, 459–482. https://doi.org/10.1002/cne.920180503

Zafeiriou, S., Zhang, C., & Zhang, Z. (2015). A survey on face detection in the wild: Past, present and future. *Computer Vision and Image Understanding*, *138*, 1–24. https://doi.org/10.1016/j.cviu.2015.03.015

Zheng, J., Li, S., & Lajoie, S. P. (2023). A review of measurements and techniques to study emotion dynamics in learning. In V. Kovanovic, R. Azevedo, D. C. Gibson, & D. lfenthaler (Eds.), *Unobtrusive observations of learning in digital environments* (pp. 7–29). Springer. https://doi.org/10.1007/978-3-031-30992-2_2

Appendix A

Differences in Performance and Emotions According to Participants' Background

Training Phases – Educational Levels

Table A-1

Flying Performance Differences According to Educational Background

Performance measure	Practice phase	Highest Degree Obtained	Mean	SD	X^2	*p*
Heading RMSE	Introduction	High school	10.11	6.75	4.46	.108
		Bachelors	7.84	3.17		
		Masters	11.83	4.93		
	Session A	High school	8.88	3.77	1.68	.432
		Bachelors	9.35	3.05		
		Masters	11.07	3.51		
	Session B	High school	10.94	0.52	1.80	.406
		Bachelors	9.37	3.51		
		Masters	11.23	4.29		
Altitude RMSE	Introduction	High school	304.42	106.16	3.80	.146
		Bachelors	239.93	92.99		
		Masters	401.78	181.05		
	Session A	High school	267.90	96.43	1.92	.383
		Bachelors	226.79	93.31		
		Masters	299.51	133.64		
	Session B	High school	227.15	32.58	3.12	.210
		Bachelors	189.55	51.13		
		Masters	267.36	94.91		
Heading expert	Introduction	High school	26.5	2.12	2.85	.241
		Bachelors	26.5	1.94		
		Masters	24.92	2.46		
	Session A	High school	32	0	6.04	.049
		Bachelors	31.22	1.39		
		Masters	28.42	3.20		
	Session B	High school	25.50	0.71	1.66	.436
		Bachelors	26.67	1.73		
		Masters	26.25	2.86		
Altitude expert	Introduction	High school	22.50	3.53	2.94	.229
		Bachelors	23.44	3.13		
		Masters	20.08	4.72		
	Session A	High school	27.00	1.41	3.53	.171

	Bachelors	28.44	2.55		
	Masters	25.00	4.37		
Session B	High school	26.00	1.41	5.29	.071
	Bachelors	26.00	1.73		
	Masters	23.00	3.76		

Note. Results based on Kruskal-Wallis tests statistic. SD=Standard Deviation.

Table A-2

SCR differences according to educational background

	Training Phases		Mean	SD	X^2	p
SCR	Introduction	High school	121.50	26.16	.896	.639
		Bachelors	105	31.53		
		Masters	107.42	24.20		
	Session A	High school	136	52.32	.153	.926
		Bachelors	126.44	32.46		
		Masters	127.08	29.21		
	Session B	High school	115.50	26.16	.694	.707
		Bachelors	117.33	30.62		
		Masters	112.08	27.44		

Note. Results based on Kruskal-Wallis tests statistic. SD=Standard Deviation.

Table A-3

Differences in Facial Expressions according to Educational Background

		Introduction				Session A				Session B			
		M	*SD*	*H*	*p*	*M*	*SD*	*H*	*p*	*M*	*SD*	*H*	*p*
NE	HS	26.50	9.19	.94	.62	34.00	16.97	1.78	.41	20.50	13.43	4.31	.12
	BA	22.00	15.1			25.12	15.73			24.62	15.86		
	Ms	34.08	26.8			34.75	16.39			41.33	23.90		
AN	HS	4.50	2.12	4.05	.13	0.5	0.7	4.16	.12	1.5	2.12	3.88	.14
	BA	34.75	20.2			29.25	13.01			24.5	16.13		
	Ms	24.16	16.2			23.25	15.23			23.41	18.35		
SU	HS	8.00	5.65	.81	.67	9.50	10.60	.87	.65	8.00	8.48	1.52	.47
	BA	15.37	25.3			14.50	22.57			16.87	21.39		
	Ms	26.33	25.5			30.08	31.46			30.25	28.59		
SA	HS	8.00	11.31	2.77	.25	5.00	7.07	4.03	.13	12.50	17.67	3.5	.17
	BA	4.12	10.1			4.25	12.08			6.37	18.03		
	Ms	7.16	9.26			12.08	17.98			13.66	16.92		
HA	HS	3.00	0	5.79	.05	14.50	14.84	2.78	.24	0.50	0.70	.88	.64
	BA	6.37	8.14			6.75	9.50			3.75	4.86		
	Ms	2.50	4.81			3.14	4.64			2.00	1.75		
SC	HS	2.50	0.70	3.22	.20	4.0	2.82	2.77	.25	0	0	5.35	.07
	BA	1.37	2.38			1.50	2.97			0.25	0.70		
	Ms	0.91	1.72			3.25	7.02			5.0	7.63		
DI	HS	0	0	2.39	.30	0	0	2.10	.35	0	0	1.15	.56
	BA	4.0	8.60			1.37	3.98			1.25	2.76		
	Ms	0.75	1.76			1.0	1.53			0.50	0.67		

Note. Results based on Kruskal-Wallis tests statistic. HS=High school, BA=Bachelor's, Ms=Master's, NE=Neutral, AN=Anger, SU=Surprise, SA=Sad, HA=Happy, SC=Scared, DI=Disgust, SD=Standard Deviation.

Participants with bachelor's degree marginally expressed more happiness than participants with a master's degree during introduction phase (Kruskal-Wallis= 2.36, p=.054).

Table A-4

Differences in Emotional Variability according to Educational Background

	Training phases		Mean	SD	KW	*p*
Emotional variability	Introduction	High school	1.89	.05	1.817	.403
		Bachelors	1.71	.25		
		Masters	1.44	.71		
	Session A	High school	1.78	.11	1.227	.541
		Bachelors	1.54	.42		
		Masters	1.62	.57		
	Session B	High school	1.24	.36	4.522	.104
		Bachelors	1.53	.35		
		Masters	1.68	.59		

Note. Results based on Kruskal-Wallis tests statistic. SD=Standard Deviation.

Training Phases – Experience with Simulators

Table A-5

Differences in Performance according to Experience with Simulators

Performance measure	Training phase	Experience with simulators	Mean	SD	U	*p*
Heading RMSE	Introduction	Yes	9.13	2.77	45	.708
		No	10.47	5.17		
	Session A	Yes	10.35	2.06	59	.609
		No	10.15	3.72		
	Session B	Yes	10.82	2.86	57	.708
		No	10.35	4.16		
Altitude RMSE	Introduction	Yes	294.74	140.57	41	.516
		No	342.42	171.08		
	Session A	Yes	235.89	114.65	43	.609
		No	271.98	114.41		
	Session B	Yes	222.10	63.83	45	.708
		No	237.40	90.44		
Heading expert	Introduction	Yes	26.67	1.96	66	.319
		No	25.35	2.37		
	Session A	Yes	30.16	2.22	53.5	.865
		No	29.70	3.09		
	Session B	Yes	27.34	1.03	67	.286
		No	26.0	2.54		
Altitude expert	Introduction	Yes	23.0	3.57	63	.431
		No	21.11	4.44		
	Session A	Yes	28.0	3.46	64	.392
		No	26.0	3.93		
	Session B	Yes	26.0	1.78	71.5	.155
		No	23.88	3.51		

Note. Results based on Mann-Whitney U tests statistic. SD=Standard Deviation.

Table A-6

Differences in SCR according to Experience with Simulators

	Training Phases	Experience with simulators	Mean	SD	X^2	*p*
SCR	Introduction	Yes	113.67	31.91	60.50	.516
		No	105.58	25.18		
	Session A	Yes	127.67	29.47	51	1.0
		No	127.58	31.99		
	Session B	Yes	119.50	28.61	57	.708
		No	112.64	27.68		

Note. Results based on Mann-Whitney U tests statistic. SD=Standard Deviation.

Table A-7

Differences in Facial Expression according to Experience with Simulators

	Training phase	Experience with simulators	Mean	SD	X^2	*p*
Neutral	Introduction	Yes	35.60	30.89	52	.493
		No	27.05	19.74		
	Session A	Yes	31	18.81	48	.704
		No	31.23	15.85		
	Session B	Yes	42.60	36.62	51	.543
		No	30.64	15.83		
Anger	Introduction	Yes	38.20	20.63	61	.164
		No	22.70	17.18		
	Session A	Yes	34	10.97	62.5	.120
		No	20.23	15.51		
	Session B	Yes	27	16.80	55	.359
		No	20.29	17.92		
Surprise	Introduction	Yes	38.40	32.48	66	.071
		No	15.47	19.99		
	Session A	Yes	44.20	32.74	66	.071
		No	16.17	23.43		
	Session B	Yes	45.80	36.44	65.50	.071
		No	16.76	17.81		
Sadness	Introduction	Yes	4.20	7.75	39.50	.820
		No	6.70	9.93		
	Session A	Yes	10	10.19	39.50	.820
		No	8.17	15.35		
	Session B	Yes	9.60	16.04	39.50	.820
		No	11.29	17.60		

Happy	Introduction	Yes	2.20	2.38	37.50	.704
		No	4.47	6.84		
	Session A	Yes	5.20	5.44	48	.704
		No	5.76	8.43		
	Session B	Yes	4.60	4.44	61.50	.140
		No	1.88	2.68		
Scared	Introduction	Yes	2	2.91	52	.493
		No	1	1.58		
	Session A	Yes	3.20	3.34	56	.319
		No	2.52	6.02		
	Session B	Yes	4.40	5.54	66.50	.058
		No	2.35	6.28		
Disgusted	Introduction	Yes	0.40	0.54	42	1.00
		No	2.29	6.10		
	Session A	Yes	0.80	1.09	48.50	.649
		No	1.11	2.84		
	Session B	Yes	0.20	0.54	42.50	1.00
		No	0.82	1.94		

Note. Results based on Mann-Whitney U tests statistic. SD=Standard Deviation.

Difficulty Levels – Educational Background

Table A-8

Flying Performance Differences According to Educational Background

Performance measure	Difficulty level	Highest Degree Obtained	Mean	SD	X^2	p
Heading RMSE	Low	High school	31.82	6.64	2.80	.246
		Bachelors	39.71	22.94		
		Masters	45.45	14.66		
	Medium	High school	51.89	20.80	2.67	.262
		Bachelors	32.55	10.74		
		Masters	41.46	16.23		
	High	High school	63.90	8.47	.911	.634
		Bachelors	69.15	16.73		
		Masters	73.93	19.58		
Altitude RMSE	Low	High school	1128.67	177.95	5.428	.066
		Bachelors	703.30	112.06		
		Masters	1072.23	500.08		
	Medium	High school	735.41	193.98	6.967	**.031**
		Bachelors	659.79	181.68		
		Masters	1137.06	489.13		
	High	High school	1869.18	627.62	0.629	.730
		Bachelors	1778.10	975.61		
		Masters	1952.27	1033.43		
Heading expert	Low	High school	20	0	3.034	.219
		Bachelors	19.66	1.0		
		Masters	19.75	1.81		
	Medium	High school	18.50	2.12	1.714	.424
		Bachelors	19.77	0.44		
		Masters	19	1.53		
	High	High school	19	1.41	.322	.851
		Bachelors	18.4	2.60		
		Masters	17.08	3.84		
Altitude expert	Low	High school	19	0	.756	.685
		Bachelors	19.11	0.78		
		Masters	17.75	2.66		
	Medium	High school	18	1.41	4.170	.124
		Bachelors	19	1.41		
		Masters	16.91	2.71		
	High	High school	16	1.41	4.587	.101
		Bachelors	16.33	3.12		

		Masters	13.33	3.11

Note. Results based on Kruskal-Wallis tests statistic. SD=Standard Deviation.

Participants with a master's degree had a higher altitude RMSE during medium difficulty tasks, compared to participants with a bachelor's degree (p=.026)

Table A-9

SCR Differences According to Educational Background

Performance measure	Difficulty level	Educational background	Mean	SD	X^2	*p*
	Low	High school	72.50	30.40	.028	.986
		Bachelors	75.33	15.71		
		Masters	74.41	20.58		
SCR	Medium	High school	83.50	34.64	.062	.970
		Bachelors	81.66	25.09		
		Masters	79.08	16.14		
	High	High school	95.50	13.43	.433	.805
		Bachelors	86.77	24.85		
		Masters	86.75	21.40		

Note. Results based on Kruskal-Wallis tests statistic. SD=Standard Deviation.

Table A-10

Differences in Facial Expressions according to Educational Background

		Low				Medium				High			
		M	*SD*	X^2	*p*	*M*	*SD*	X^2	*p*	*M*	*SD*	X^2	*p*
NE	HS	21	9.89	0.92	.62	13	1.41	3.54	.17	20.5	4.94	7.74	.021
	BA	15.8	15.8			18.2	11.71			15.6	9.42		
	Ms	21.4	14.2			27.3	16.5			27.3	9.55		
AN	HS	0	0	6.59	.03	1	1.41	3.17	.20	1	1.41	4.02	.134
	BA	17.7	6.96			17.6	13.8			18.3	9.25		
	Ms	12.3	9.91			16	12.9			18.3	11.68		
SU	HS	6.50	9.19	1.07	.58	5.50	4.94	.780	.67	5.50	4.94	2.75	.252
	BA	11.37	13.5			10.3	13.7			9.62	13.7		
	Ms	17.0	15.2			20.8	19.6			22.4	19.6		
SA	HS	5	7.07	3.89	.14	3.50	4.94	4.74	.09	9	12.7	2.39	.302
	BA	3.12	8.83			5.37	12.3			3.12	8.83		
	Ms	7.75	9.78			9.25	11.74			8.75	13.9		
HA	HS	5.50	3.53	3.63	.16	2	3.53	.077	.96	2	2.82	1.49	.474
	BA	1.75	2.86			1.50	2.86			1.50	2.32		
	Ms	1.16	1.33			1.66	1.33			1.66	2.14		
SC	HS	1.50	4.73	1.72	.42	0.50	0.70	1.40	.49	0.50	1.41	7.33	.026
	BA	1.37	0.70			0.37	1.06			0.37	0		
	Ms	2.16	2.66			1.91	3.31			1.91	3.31		
DI	HS	0	0	1.21	.54	0	0	.948	.62	0	0	1.34	.511
	BA	0.25	0.46			0.12	0.35			0.62	1.40		
	Ms	0.58	0.90			0.25	0.45			0.66	1.15		

Note. Results based on Kruskal-Wallis tests statistic. HS=High school, BA=Bachelors, Ms=Masters, SD=Standard Deviation, NE=neutral, AN=anger, SU=surprise, HA=Happy, SC= Scared, DI= Disgust.

Participants with master's degree expressed more neutral than participants with a bachelor's degree during high-difficulty tasks (*p*=.019).

Participants with bachelor's degree expressed more anger than participants with a master's degree during low-difficulty tasks (*p*=.036).

Table A-11

Emotional Variability according to Educational Background

	Difficulty level	Educational background	Mean	SD	X^2	*p*
Emotional variability	Low	High school	1.59	.06	3.480	.176
		Bachelors	1.53	.25		
		Masters	1.79	.33		
	Medium	High school	1.73	0.01	2.576	.276
		Bachelors	1.38	0.47		
		Masters	1.53	0.57		
	High	High school	1.55	.39	1.583	.453
		Bachelors	1.77	.35		
		Masters	1.68	.36		

Note. Results based on Kruskal-Wallis tests statistic. SD=Standard Deviation.

Difficulty Levels – Educational Background

Table A-12

Performance Differences according to Experience with Simulators

Performance measure	Difficulty level	Experience with simulators	Mean	SD	U	*p*
Heading RMSE	Low	Yes	44.76	17.01	61	.516
		No	41.05	18.52		
	Medium	Yes	34.87	11.75	45	.708
		No	40.30	16.22		
	High	Yes	79	12.71	72	.155
		No	67.90	18.46		
Altitude RMSE	Low	Yes	750.49	216.63	31	.177
		No	997.11	444.60		
	Medium	Yes	814.31	282.41	48	.865
		No	951.04	480.51		
	High	Yes	1877.11	1075.27	55	.812
		No	1876.81	939.79		
Heading expert	Low	Yes	19.16	2.04	55	.812
		No	19.23	1.34		
	Medium	Yes	19.50	0.83	53.50	.865
		No	19.17	1.42		
	High	Yes	18.83	1.83	63.50	.392
		No	17.42	3.58		
Altitude expert	Low	Yes	19.50	0.54	76.50	.074
		No	18	2.26		
	Medium	Yes	18.83	1.47	65.60	.319
		No	17.47	2.52		
	High	Yes	16.66	3.44	63.50	.392
		No	14.41	3.26		

Note. Results based on Mann-Whitney U tests statistic. SD=Standard Deviation.

Table A-13

SCR Differences according to Experience with Simulators

	Difficulty level	Experience with simulators	Mean	SD	U	*p*
SCR	Low	Yes	74.66	19.97	54.50	.812
		No	74.58	18.67		
	Medium	Yes	87.33	24.82	63	.431
		No	78.05	18.87		
	High	Yes	85.16	17.74	45.50	.708
		No	87.58	23.01		

Note. Results based on Mann-Whitney U tests statistic. SD=Standard Deviation.

Table A-14

Facial Expressions Differences according to Experience with Simulators

Emotion	Difficulty level	Experience with simulators	Mean	SD	U	*p*
Neutral	Low	Yes	20.20	16.46	44	.940
		No	19.11	11.44		
	Medium	Yes	28.88	24.19	52	.493
		No	20.94	11.18		
	High	Yes	24.60	15.22	54.50	.359
		No	21.82	9.22		
Anger	Low	Yes	17.40	8.14	56	.319
		No	11.94	9.88		
	Medium	Yes	23.20	13.86	61	.164
		No	12.88	12.38		
	High	Yes	20.40	6.80	52.50	.446
		No	15.70	12.20		
Surprise	Low	Yes	24.40	15.69	66.50	.058
		No	11	13.41		
	Medium	Yes	34.80	29.88	66	.071
		No	10	12.54		
	High	Yes	30.80	22.43	64	.101
		No	11.94	14.25		
Sadness	Low	Yes	4.20	7.36	39.50	.820
		No	6.29	9.74		
	Medium	Yes	5.60	10.92	36	.649
		No	7.35	11.90		
	High	Yes	9.80	17.78	45.50	.820
		No	5.82	10.27		

Happy	Low	Yes	3	3.16	58.50	.218
		No	1.41	2.12		
	Medium	Yes	2.20	1.30	62.50	.120
		No	1.47	2.34		
	High	Yes	4.60	6.30	43.50	.940
		No	3.52	4.37		
Scared	Low	Yes	3.20	2.77	64	.101
		No	1.41	3.27		
	Medium	Yes	2	2.54	58.50	.218
		No	1	2.64		
	High	Yes	2.40	3.28	56.50	.283
		No	1.17	2.50		
Disgusted	Low	Yes	1	1	62	.140
		No	0.23	0.56		
	Medium	Yes	0.20	0.44	43.50	.940
		No	0.17	0.39		
	High	Yes	0	0	25	.189
		No	0.76	1.30		

Note. Results based on Mann-Whitney U tests statistic. SD=Standard Deviation.

Table A-15

Emotional Variability Differences according to Experience with Simulators

	Difficulty level	Experience with simulators	Mean	SD	U	*p*
Emotional variability	Low	Yes	1.84	0.32	59.50	.189
		No	1.63	0.30		
	Medium	Yes	1.59	0.37	43	1.0
		No	1.51	0.55		
	High	Yes	1.68	0.45	42	1.0
		No	1.68	0.34		

Note. Results based on Mann-Whitney U tests statistic. SD=Standard Deviation.

Bridging Text

The results of Chapter 3 demonstrate that beginner pilot trainees' performance and emotions change according to training phase and task difficulty. Pilot trainees were more accurate in the final training phase, and during low and medium difficulty tasks. Moreover, trainees had more expression of neutral and anger, and it is argued that the expression of these emotions might suggest that they were concentrated on the task. Interestingly, trainees experienced more skin conductance responses and fluctuations across multiple emotions (i.e., emotional variability) during the middle training phase and during high-difficulty tasks. The results of this study contribute to understanding the role of a wider range of emotional responses and emotion dynamics in flight training, which might inform instructors of key moments of emotional arousal in relation to performance accuracy.

The findings of Chapter 3 demonstrate that emotion dynamics vary during a simulated flying task. However, the relationship between emotional variability and flying performance was not explored directly. Moreover, following the control-value theory and studies found in Chapter 2, a factor that might explain the relationship between affect and flying performance are cognitive appraisals. Thus, the objective of Chapter 4 is to delve into the relationship between emotional variability and flying performance as moderated by perceived control and value over the task, as trainees transition across the different training phases. This chapter attempts to link aviation and educational theories exploring emotions by guiding the research using Pekrun's (2019) control-value theory of achievement emotions in the context of flight training.

Chapter 4. Manuscript 3

Exploring Pilot Trainees' Perceived Control and Value, Emotional Variability, and Flying Performance

Ruiz-Segura, A., Law, A., Jennings, S., Bourgon, A., Churchill, E., & Lajoie, S. (under review). A study of pilot trainees' perceptions, emotional variability, and performance. *Aerospace Medicine and Human Performance*.

Abstract

This study explores the relationship between emotional variability and performance during flight simulation training, particularly exploring whether trainees' perceived control and value explain when this relationship is meaningful. Pilots' perceptions and affective responses affect cognition and performance in flight training. Perceived control and value refer to appraisals of agency and importance of the task that work as precursors of emotions in achievement situations. Emotional variability is studied as fluctuations of multiple emotions over time, and it is expected to explain the impact of emotions on performance accuracy. Twenty-two volunteers performed basic flight maneuvers. The experiment consisted of three phases: introduction (seven tasks), session A (eight tasks), and session B (seven tasks). Flying performance was measured using root-mean-square-error and expert ratings. Emotions were inferred from video-recordings of facial expressions and quantified to calculate emotional variability. Participants reported control and value when they finished each phase. Findings confirm that pilot trainees' perceived control and value over the task moderate the relationship between emotional variability and performance in a simulated flying task. These interactions varied across different training phases: perceived value was a significant moderator during introduction, and perceived control was a significant moderator during sessions A and B. Participants who perceived low value or low control and displayed low emotional variability performed poorly, while those with high emotional variability performed more accurately. Emotional variability influences flying performance, moderated by perceived control and value. Emotional variability might be an adaptive response, particularly, when trainees report low agency and importance over the task.

Keywords: Affective factors, Emotion, Simulations, Simulation and training, Learning

Introduction

The aviation industry uses simulations as a fundamental training technique for training in safe and authentic environments (Hamman, 2004). Recent evidence demonstrates that pilots' emotion and cognition affects flying accuracy (Herrera-Aliaga & Estrada, 2022). Yet, like other high-stakes professions, the culture in aviation might encourage pilot trainees to disconnect from their emotions while flying (Duffy et al., 2016). Psychology research shows that uncontrolled intense emotions and low perception of ownership over the task might negatively affect performance (Bailen et al., 2019; Pekrun, 2019). This study attempts to explore the relationship of pilots' emotional experience and flying performance during a simulated flying task. Particularly, this study examines the moderating role of subjective perception of control and value over the task to explain the relationship between emotional variability and flying performance.

Perceptions, Affect, and Flying Performance

Pilots' perceptions and affective reactions influence cognition and decision making, having consequences in flight accuracy (Gaetan et al., 2015). In fact, the theories used to explain affective responses in flight training significantly rely on understanding pilot's perception of importance and control over the task. Sandra Hart, creator of NASA-TLX, was a pioneer exploring the connections between pilots' perceptions, stress, and performance (2006). A preliminary study to design the NASA-TLX consisted of asking pilots their perception over stereotypical segments of flights, finding that pilots perceived that challenging flight segments and errors increased stress, hampering subsequent performance (Hart & Bortolussi, 1984). The model guiding NASA-TLX focuses mainly on workload, accounting for mental resources (e.g., stress) and perceptions as secondary elements. This pattern is frequently encountered in flight training research, in which affective responses, such as stress and emotions, are a factor influencing cognition, having consequences on performance.

For instance, the impact of anxiety on cognition and performance has been explained using attentional control theory (Eysenck et al., 2007), which states that individuals achieve similar accuracy levels, despite anxiety levels. However, high anxiety requires more mental resources, causing less attentional control to solve the task efficiently (Eysenck et al., 2007). Studies using this theory found that performance in simulated flight is similar in neutral or anxiety-inducing conditions, yet anxious conditions result in poorer performance efficiency (i.e., more dwell time and root-mean-square error) (Allsop & Gray, 2014; Gray et al., 2016). Another study explored the ratio between pilots' perception of task demands (negative) and personal resources (positive), showing that the difference between both perceptions positively related to flying performance including instructors' evaluation, and deviations from ideal flight (Vine et al., 2015). These results demonstrate a relationship between task perception and personal capacities regarding flying performance (Vine et al., 2015); however, the interpretation of these results is unclear, leaving ambiguous interpretations of the impact of trainees' perceptions on flying accuracy.

Hart's workload and human performance framework (Hart, 2006; Hart & Bortolussi, 1984), along with the attentional control theory (Eysenck et al., 2007) show that stress and anxiety affect flying performance by influencing parallel cognitive processes. Still, a limitation is of these models is that they only discuss displeasing and physiologically arousing states, disregarding other affective responses (Gross, 2015). For that reason, this study expands our understanding of emotions in aviation using the control value theory of achievement emotions (Pekrun, 2019).

It should be noted that attentional control and perceived control refer to two different concepts. Attentional control refers to the trainees capacity to concentrate on the task to achieve one's goal, despite having distracting stimuli (Eysenck et al., 2007), whereas perceived control refers to the subjective evaluation of one's ability to achieve a goal (Pekrun, 2019). Similarly, although cognition, affect, and performance are connected, they are independent terms.

Achievement Emotions

Emotions are short affective states triggered by a specific stimuli, lasting from seconds to minutes (Gross, 2015). Emotions are classified according to valence, the subjective perception of pleasantness, and arousal, the associated physiological response (Gross, 2015; Pekrun, 2019). Emotions differ from other affective states like stress and mood. Stress can occur with unclear triggers during challenging situations, above one's capabilities, and is experienced as displeasing, with increases in physiological arousal (Gross, 2015). Moods have low intensity, are less specific (like "feeling down") and are longer, sustaining from hours to days (Gross, 2015). This study focuses on emotions to understand affective states triggered by flying simulations.

The control-value theory (CVT) explores emotions occurring in situations attached to success and failure (i.e., achievement situations) (Pekrun, 2019). CVT assumes that control and value appraisals will determine emotions that learners experience (Pekrun, 2019). Perceived control refers to the degree of agency that learners has over the task, signaling whether success associates to oneself or to external factors (Pekrun & Perry, 2014). Perceived value indicates evaluation of the task as supporting or blocking one's goals (Pekrun & Perry, 2014). High perceived control in oneself might change depending on the value over the task: if the outcome is positive (aligned with one's goals), learners will experience joy, whereas if the outcome is negative (mismatching personal objectives), learners will experience anger (Pekrun & Perry, 2014). Emotions can occur at different timepoints of the activity, prospectively, during, or retrospectively (Pekrun & Perry, 2014). In this study, novice trainees can perceive achieving or failing to learn flying maneuvers as an achievement situation to become pilots (T. Li & Lajoie, 2021). This study explores the following activity emotions: joy (positive-activating); anger, disgust, and fear (negative-activating); sadness (negative-deactivating); surprise (non-valenced and activating); with neutral (non-valenced) as a baseline affective state (Pekrun, 2019).

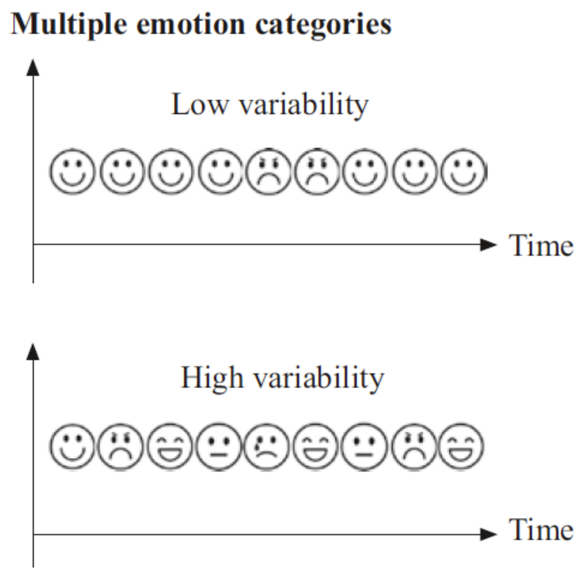
A previous study found that pilot trainees' perceived control negatively correlated with experiencing fear and perceived value negatively related to emotional arousal, inferred from facial expressions (T. Li & Lajoie, 2021). To our knowledge, only this study has explored control and value appraisals in flight training. Yet, control and value appraisals and performance accuracy have been examined in other professional domains. Duffy et al. (2018) explored achievement emotions when learning with authentic surgical simulations, including a supervised surgery, a computer-based simulation, and a mobile application. In the three scenarios perceived control positively correlated with more accurate performance, positive-activating and positive-deactivating emotions, and negatively correlated with negative-deactivating emotions (Duffy et al., 2018). Value only correlated with positive-deactivating emotions when using the mobile app (Duffy et al., 2018). In this study we attempt go beyond discrete emotions by exploring emotional dynamics when learning flying maneuvers in a simulator. This study will explore emotion dynamics acknowledging and understanding the time-dynamic nature of emotions, in relation to trainees' evaluation of control and value over the task. By exploring emotion dynamics, we add a layer recognizing the stability or changes in one emotional feature (i.e., intensity) and across multiple emotions (i.e., emotional variability) (Kuppens, 2015).

Emotional Variability

When studying emotions and performance, emotions are traditionally analyzed based on frequency and duration (Kuppens, 2015; S. Li et al, 2021a). New approaches propose viewing emotions as a dynamic continuum throughout training (D'Mello & Graesser, 2012; Zheng et al., 2023a). Emotions dynamics refer to patterns and regularities characterizing changes and fluctuations in emotions over time, their underlying processes and consequences (Houben et al., 2015; Kuppens, 2015). Emotion dynamics can be analyzed as changes in features like frequency, intensity, and variability in one or more emotional subcomponents (Bailen et al., 2019; Kuppens, 2015). Frequency refers to how often an emotion occurs (Bailen et al., 2019). Intensity refers to the strength of single emotional responses, i.e., strength of anger (Bailen et al., 2019; Zheng et al., 2023a). Emotional variability refers to fluctuations in multiple emotions over time. As exposed later in the method sections, in this manuscript we only counted the frequency of emotions that were intense. We only counted the frequency of emotions that were sufficiently strong to be recorded in the system and higher than the rest of the emotions (Zheng et al., 2023). Later, we used the frequency of these intense emotions to fluctuations among basic emotions, like anger, surprise, joy, etc. (S. Li et al., 2021a). Figure 20 shows a visual representation of emotional variability, pertinent to this study.

Figure 20

Visual Representation of Emotional Variability



Note. Adapted from “A review of measurements and techniques to study emotion dynamics in learning” by J. Zheng, S. Li, and S. Lajoie, in V. Kovanovic, R. Azevedo., D.C. Gibson, and D. Ifenthaler (Eds), *Unobtrusive Observations of Learning in Digital Environments* (p. 10), 2023, Springer. Copyright 2023 Springer. Reprinted with permission.

In simple words, emotional variability is seen as the “ups and downs” in emotions, compromising intense changes across positive and negative emotions (Bailen et al., 2019). Interestingly, emotional variability has been mainly studied as a psychopathological symptom. Greater emotional variability is associated to less favourable outcomes, despite positive and negative emotions levels, including emotional dysregulation and depressive symptoms in neurodivergent people (Gruber et al., 2013). Recent studies are exploring adaptive (or maladaptive) functions of emotional variability in short-term everyday situations (Houben et al., 2015; Xu et al., 2016). For example, a study found that emotional variability positively related to job dissatisfaction, hypothesizing that workers with higher emotional variability invest more efforts in regulating emotional variations, resulting in more emotional fatigue and more job dissatisfaction (Xu et al., 2016).

In professional training, medical students' emotional variability related to performance diagnosing simulated virtual patients (S. Li, et al., 2021a). Descriptive patterns suggest that students with better performance had less emotional variability compared to those with less accurate performance (S. Li, et al., 2021b). To our knowledge, only Gaetan et al. (2015) discussed potential implications of emotional variability in pilot training, noting that novice pilots experience higher emotional variability than intermediate and expert pilots; however the measurement of emotional variability remains unclear. Studying emotional variability in non-psychopathological scenarios is fairly recent; therefore, there remains a lot unknown about the emotional variability impact on training outcomes (Kashdan & Rottenberg, 2010; S. Li et al., 2021a). Following Pekrun's CVT (2019), this study argues that the relationship between emotional variability and performance will be further influenced by trainees' perceived control and value over the task.

Current Study

This study examines the connection between emotional variability and performance in pilot training using simulated flying tasks, specifically exploring how pilot trainees' perceived control and value moderates this relationship. The central research question is: Do perceived control and value moderate the relationship between participants emotional variability and flying performance?

Hypotheses

Perceived control and value over the task will moderate the relationship between emotional variability and flying performance (Pekrun, 2019). A previous study conducted by the authors that identified significant effects of training on flying performance and emotions, thus the hypotheses and analyses are conducted in alignment with training phases (introduction, session A, and session B):

1. Perceived Control:
    a. High control with low emotional variability is expected to associate with more accurate flying performance across all training phases (Duffy et al., 2020; S. Li, 2021a; S. Li, et al., 2021b).
    b. During introduction phase, an interaction term is expected for low control: low control with high emotional variability is likely correlated with poorer performance, while low control with low emotional variability will relate to better performance (Duffy et al., 2020; S. Li, 2021a; S. Li, et al., 2021b).
    c. During session B, an opposite effect is expected: low control with higher emotional variability is expected to relate to better performance than low control with low emotional variability. Specifically, lower emotional variability might indicate lower emotional engagement toward the end of the task (D'Mello & Graesser, 2012). Session A will be examined using an exploratory approach.
2. Perceived value:
    a. High task value with low emotional variability is expected to correlate with more accurate performance across all the training phases (Duffy et al., 2020; S. Li, 2021a; S. Li, et al., 2021b).

b. Low perceived value with low emotional variability will relate to poorer performance across training phases, as participants may lack motivation as tasks progress (Pekrun, 2019).

c. The relationship between higher emotional variability with high perceived value depends on the training phase. During introduction phase, high emotional variability with high value is expected to relate to poorer performance since more emotional fluctuations might imply less emotional control, seen as a barrier for goal achievement (Bailen et al., 2019).

d. During session B, the same pattern might relate to better performance as more emotional variability might be associated with greater emotional engagement, facilitating goal achievement (D'Mello & Graesser, 2012). An exploratory approach will be taken regarding low perceived value with high emotional variability since low perceived value generally relates to reduced task effort in the task (Pekrun & Linnenbrink-Garcia, 2014). It is unclear whether emotional variability will act as a barrier or as a stimulant toward effective performance.

Methods

Participants

This study was part of a larger project that attempted to understand behavioral and biometric changes of ab-initio pilot trainees, i.e., trainees with little to no flying experience (Marques et al., 2023). The requirements to apply to aviation school in North America are (1) being 14 years or older; (2) being able to understand, speak, read, and write English; and (3) passing a medical examination (FAA Department of Transportation, 2003; Transport Canada, 2019). Volunteers were recruited from a large North American city with the beforementioned criteria. As the larger study required a significant time commitment and commuting, participants were recruited by convenience in awareness that they had the time flexibility to participate in the study.

Following ethics approval, twenty-two volunteers participated in the study (M_{age}=28.77, SD_{age}=4.71), 12 identified as females (54.54%) and 10 as males (45.45%). Participants reported their most recent educational degree, including high-school (n=2), bachelor's (n=8), and master's (n=12) diplomas. Five participants had prior experience using flight simulators, four of them reported having use a flight simulation one time, and one did not specify. Participants reported to be free from medical conditions limiting their participation. Only relevant procedures are presented. A power analysis for multiple regressions revealed that n=25 would be sufficient, setting Cohen's f=.35, α=0.05, power 1-β=0.80, expecting a large effect size in G*Power software (Faul et al., 2007).

Apparatus

The experiment consisted of training study volunteers to simulate ab-initio pilot training, performing flying twizzles in a fixed-based simulator designed and operated by Marinvent Corporation. Ab-initio training focuses on developing foundational knowledge and skills for pilots with little to no experience. The cockpit included a control yoke, throttle, and pedals, and a screen showing an aircraft primary flight display. The throttle and pedals were operated on autopilot to help reduce workload level. Participants controlled the aircraft turn rate and heading control by using the yoke to control aircraft roll, turning the yoke left and right. Aircraft pitch, used to set climb/descent rates and thus control altitude, was controlled by moving the yoke forward and backwards. X-Plane 11, a simulation software package designed to reflect real the behaviour of real aircrafts (Laminar Research, 2022), was used to render the primary flight display. A camera on top of the primary flight display was used to record participant's facial expression, and a tablet was on the yoke's left for providing instructions and questionnaires.

Flying tasks

Twizzles (basic maneuvers) were designed by experienced instructors. Participants were instructed to change heading (i.e., turns) and altitude (climbs and descends) without outside visual cues and relying on instrument indication. Twizzles had three difficulty levels: low, medium, and high, see Table 13 for sample instructions.

Table 13

Example of Instructions for Maneuvers according to Difficulty Levels

Difficulty level	Instructions
Baseline	Maintain straight and level at 10 000 feet altitude, and a heading of 0°
Easy	Maintain heading at 0 degrees. At the same time: CLIMB to an altitude of 11,000 feet at 1000 fpm and level off.
Medium	Maintain heading at 0 degrees. At the same time: CLIMB to an altitude of 10500 feet at 1000 fpm; Then DESCEND back to an altitude of 10,000 feet at 1000 fpm and level off.
Difficult	Turn LEFT at 30 deg AOB to a heading of 240 degrees and roll out on a steady heading. At the same time: CLIMB to an altitude of 10500 feet at 1000 fpm; Then DESCEND back to an altitude of 10,000 feet at 1000 fpm and level off.

Procedures

This study is part of a larger project, only relevant procedures are presented in Table 14. Participants were briefed on the experiment's objectives and provided consent. Participants completed questionnaires about demographics and relevant previous experience. Participants watched video-training and received instructions using the cockpit and performing flying maneuvers. The experiment consisted of three phases: introduction (seven tasks), session A (eight tasks), and session B (seven tasks), totalling 22 tasks. Each task had a 30-second baseline and a 90-seconds trial. X-plane metrics were re-set for each task to start from a 10,000 feet altitude, 247 Kts (controlled) speed, and 0° bank angle. Introduction tasks were identical for all participants, with increasing difficulty, and participants received feedback from a trained researcher to ensure their understanding. In sessions A and B, the difficulty order was randomized, with five tasks for each difficulty level. Control-value questionnaires were completed after tasks 4, 6, and 7 (averaged for analyzing introduction phase) and at the end of sessions A (after task 15) and B (after task 22). Breaks were provided between phases, resulting in a six-hour experiment.

Table 14

Sample Procedures of Experiment

Sign consent	
Demographics	
Video-training	
Guided hands-on familiarization	
Introduction	
1.	Low
2.	Low
3.	Low
4.	Low – CVQ
5.	Medium
6.	Medium – CVQ
7.	High
	CVQ

Session A	
8.	Medium
9.	Low
10.	High
11.	Low
12.	High
13.	Medium
14.	Low
15.	High
	CVQ

Session B	
16.	Medium
17.	Low
18.	High
19.	Medium
20.	High
21.	Medium
22.	Low
	CVQ

Note: CVQ= Control-value questionnaire

Measures

Flying performance

Flying performance was evaluated using aircraft state data from X-Plane logfiles. Heading and altitude were selected to assess pitch and roll axes.

Flying error. Root-mean square error (RMSE) was calculated by comparing actual heading or altitude from target values at each time point, squaring the error, and computing the square root of mean error. Since the maneuvers are continuous and some require reversal paths, sinusoidal target functions were used to connect start, middle, and end points for instructed heading and altitude across time frame for each twizzle, accounting for expected continuous changes for an idealized flight path (Jennings et al., 2024). A larger RMSE reflects more deviation from the target path, thus more error.

Expert rating. An aviation instructor rated participants' performance based on graphs visualizing flying trajectories. Scores ranged from 1 to 4, with 1 indicating very low accuracy, and 4 representing high accuracy compared to the instructed metric, 0 represented a non-completed task. A higher score represents higher performance.

This rating is standardly used in the industry partner supporting the larger project; however, the rating is usually performed with the instructor directly observing the student pilot in the simulation. This new rating, visualizing graphs with the flight path is a novel approach. Preliminary analysis showed that expert rating observing graphs aligned with objective measures of performance, and more analyses are being performed to assess its reliability (Jennings et al., 2024).

Emotional Variability based on Facial Expressions

Participant facial expressions were recorded using a camera mounted atop primary the primary flight display screen. Those using corrective lenses wore contact lenses for better facial expression identification. Video-analysis was performed using FaceReader 6.0, a software application trained to automatically detect facial expressions and analyze emotions (Loijens et al., 2015). FaceReader analyzes facial expressions, resulting in detecting emotions, following three main stages (Loijens et al., 2015). First, FaceReader uses a deep-learning algorithm to detect the face (Viola & Jones, 2004; Zafeiriou et al., 2015). In the second stage 500 key points of the face are mapped using an active appearance model, a deep neural network (Cootes & Taylor, 2001; Loijens et al., 2015). The key points of the face are integrated using a principal component analysis resulting in a vector, which is later entered as the input of an artificial neural network (ANN) to identify the emotional qualities of the facial expression (Loijens et al., 2015). The ANN was trained manually coding 10,000 images using the facial action coding system (Ekman & Rosenberg, 2005).

The results of the analysis result in providing information about diverse emotional information inferred from the facial expression (Loijens et al., 2015). This manuscript uses the detection of seven basic emotions (anger, disgust, happy, neutral, sad, scared, and surprised, with 90% accuracy (Loijens et al., 2015). To control for individual differences, before starting the analysis FaceReader was calibrated by selecting entering a baseline video showcasing a representative neutral expression of each participant. The sample rate was set for 30 samples per second.

The output of FaceReader measures the intensity of emotions by quantifying emotions between 0 and 1, in which 0 implies that the emotion is not present, and 1 indicating that the emotion is fully present according to the ANN (Loijens et al., 2015). For this manuscript, the state log was used, which records the time stamp and label of dominant emotions. A dominant emotion is recorded when (1) its intensity is higher than all other emotions, and (2) the emotion is sustained for more than 0.5 seconds (Loijens et al., 2015). Therefore, we argue that dominant emotions inherently are intense.

The frequency of dominant emotions was used to calculate emotional variability. Emotion variability encompasses fluctuations in emotional states, offering dynamic insight beyond mere emotions frequency (S. Li, et al., 2021a). Shannon's entropy (1948) formula was applied to assess the randomness of emotional states (Jack et al., 2014; S. Li, et al., 2021a).

As the duration of the task was equal for all participants, we counted the raw frequency of dominant emotions per training phase.

$$h(p_1,\ldots, p_a) = -\sum_{i=1}^{a} pi\log_2(p_i)$$

Where p_i represents the probability of emotion *I* occurring in a sequence of emotions. In this case, the seven basic emotions that were detected during the task. Using the binary logarithm ($\log_2$) identifies the times the number needs to be multiplied by itself to obtain *pi.* Consequently, the sum of the $\log_2$ of the probabilities of the seven categories of emotions would show the degree of randomness (Karaca & Moonis, 2022). The minimal value of entropy is zero, indicating that the person expressed only one dominant emotion throughout the task, having the highest probability of occurring. The maximum value is 2.8 ($\log_2(7)$), shows that the seven emotions were equally present, having an equal probability, and thus the highest emotional variability (Zheng et al., 2023a).

Following this definition, having a value of zero emotional variability does not necessarily imply neutrality. Rather, having a low emotional variability implies that the person expressed only one dominant emotion, which could be any on the seven emotions.

Control and Value Questionnaires

Appraisals were measured specific to flying tasks (T. Li & Lajoie, 2021). Questionnaires employed five-point Likert scales ranging from “strongly disagree” to “strongly agree”, quantified from 1-5, transcribed and completed using Qualtrics (Copyright © 2020, Provo, UT). Three items were adapted from the academic control scale, scores of the items were averaged for analysis (Perry et al., 2001, Cronbach’s α=.80). A sample item is “I feel the more effort I put in, the better I did at this flying task”. Due to the complexity of the experiment, it was jointly decided with the partners to reduce the amount of self-reported items. Consequently, it was decided to adapt and include only three out of the eight items from the original scale.

Value appraisals were gauged through five adapted items from the importance, usefulness, and interest of the expectancy-value questionnaire (Wigfield & Eccles, 2000), an average score was used for analysis (Cronbach’s α importance=.79, usefulness=.81, and interest=.79 , Gao & Xiang, 2008). A sample item is “I think I can use what I learned from doing this flying task to other things in other situations”. The full list of items can be found in Appendix B.

Due to the heterogeneity of participants, nonparametric tests were conducted to identify potential differences according to highest educational degree obtained and previous experience with flight simulators, see Appendix C. Participants with a bachelor's degree reported a higher degree of perceived control during the introduction, compared to participants with a master’s degree (X^2= 9.43, p=.004). Performance metrics and emotional variability did not differ regardless of educational level, nor previous experience with flight simulations.

Results

The relationship between emotional variability and flying performance, moderated by control and value appraisals, was investigated through moderation analyses. Beyond identifying correlations, moderation analysis shows when or under what circumstances X exerts an effect on Y (Hayes, 2022). In this study we attempt to identify what levels of perceived control and value (W) contribute to explain the relationship between emotional variability (X) and flying performance (Y). IBM® SPSS® version 29 and PROCESS macro were employed to compute a moderation model, accounting for multiple regression models following the equation Y=*i*+X+W+XW+e (Haynes, 2018): Y represents flying performance metrics, *I* is the constant, X is emotional variability, W is perceived control or value, XW is the interaction between emotional variability and perceived control or value, and e denotes error. Johnson-Neyman technique was applied to identify significant moderation limits, different from zero. Graphs were created using CAHOST Excel workbook (Carden et al., 2017). Descriptive statistics are summarized in Table 15. Moderation analyses using heading performance, including both RMSE and expert rating, did not yield statistically significant results.

A previous study confirmed that performance metrics significantly differed across training phases, showing that performance became more accurate as the training phases advanced. Therefore, a repeated-measures ANOVA was conducted to identify differences in perceived control and value across training phases. Results showed that perceived control and value were significantly higher after session B and marginally higher after session A ($p_{control}$=.02, p_{value}=.035) compared to introduction phase.

Table 15

Descriptive Statistics and Repeated Measures ANOVA for Perceived Control and Value

	Introduction		Session A		Session B		RM-ANOVA	
	M	*SD*	*M*	*SD*	*M*	*SD*	*F(2,21)*	*MD*
Emotional variability	1.64	0.44	1.60	0.48	1.59	0.49		
Control	3.67	0.45	3.95	0.60	4.13	0.61	10.35**	I<SB**
Value	3.86	0.41	3.95	0.44	4.04	0.51	7.46**	I<SB**
Heading RMSE	10.27	4.70	10.22	3.40	10.37	3.86		
Altitude RMSE	335.6	163.4	265.6	114.7	236.9	83.4		
Heading rating	3.66	0.33	3.71	0.36	3.75	0.33		
Altitude rating	3.05	0.59	3.29	0.48	3.47	0.47		

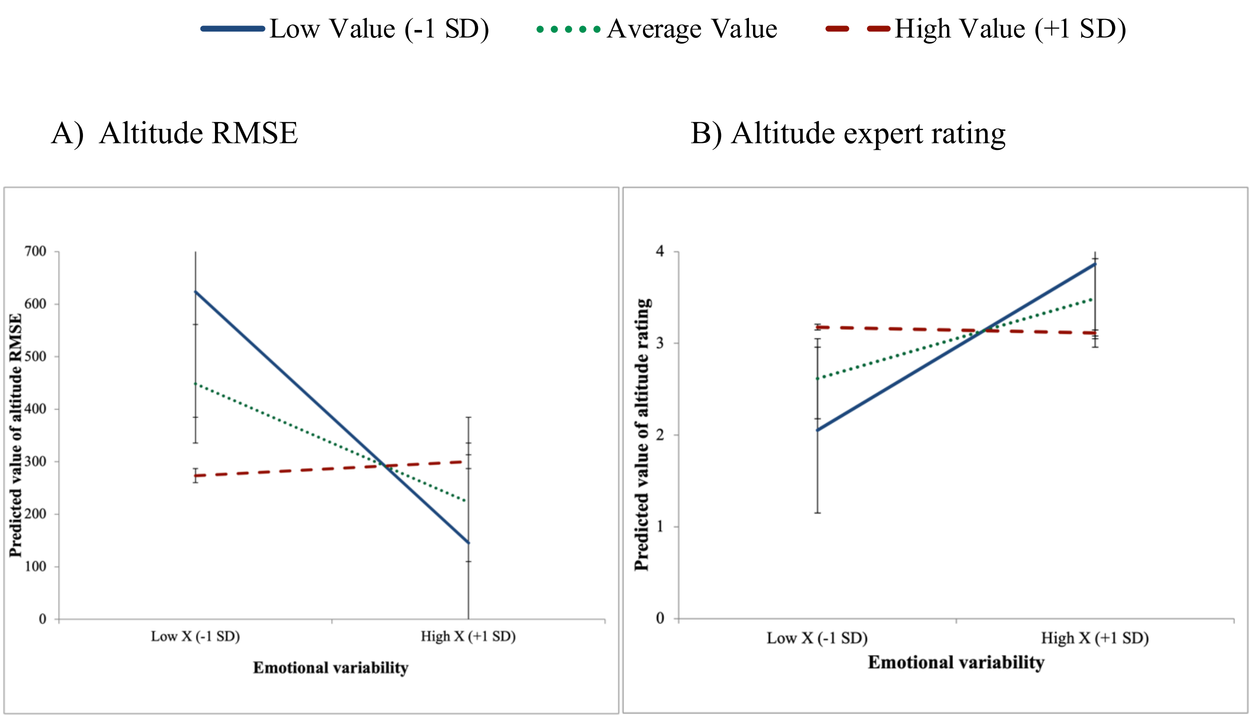
Note. M=Mean, SD=Standard deviation, MD=Mean difference, I=Introduction, SB=Session B. **p<.001, *<.017 after Bonferroni correction

Introduction Phase

During introduction phase, models employing control did not yield statistically significant results. Perceived value moderated the relationship between emotional variability and altitude performance (altitude RMSE R^2=.378, F(3, 18)=3.65, p=.03, altitude rating R^2=.389, F(3, 18)=3.83, p=.03). Emotional variability independently contributed to explain altitude performance variance (RMSE R^2=.418, t=-3.15, p=.005; expert rating R^2=.438, t=3.38, p=.04). The interaction term, emotional variability by perceived value, marginally explained altitude RMSE variance (R^2=.112, F=3.23, p=.08) and expert rating (R^2=.116, F=3.43, p=.08). Johnson-Neyman indicated significant regions for the association between emotional variability and altitude performance at a mean level ($p_{altitude\ RMSE}$=.005, $p_{altitude\ rating}$=.003) and -1SD of perceived value ($p_{altitude\ RMSE}$=.02, $p_{altitude\ rating}$=02). Simple slope analyses (Figure 21) demonstrated that participants who perceive medium or low value and exhibited lower emotional variability had less accurate performance, characterized by larger RMSE and lower ratings. Conversely, participants who perceived mid to low value and displayed high emotional variability, had more accurate performance. Descriptively, it is observed that with lower emotional variability, perceived value has a greater effect on flying performance.

Figure 21

Moderating Effect of Perceived Value Between Emotional variability and Altitude Performance during Introduction Phase



Note. Error bars correspond to standard errors of performance metrics.

Session A

In session A, models assessing the moderating role of perceived value did not yield statistically significant results. Perceived control moderated the relationship between emotional variability and altitude performance (RMSE R^2=.425, F(3, 18)=4.43, p=.02; expert rating R^2=.340, F(3, 18)=3.09, p=.05). Perceived control independently contributed to explaining altitude rating variance (R^2=.394, t=2.34, p=.03). The interaction term emotional variability by perceived control accounted for additional variation in altitude RMSE (R^2=.259, F(1, 18)=8.11,

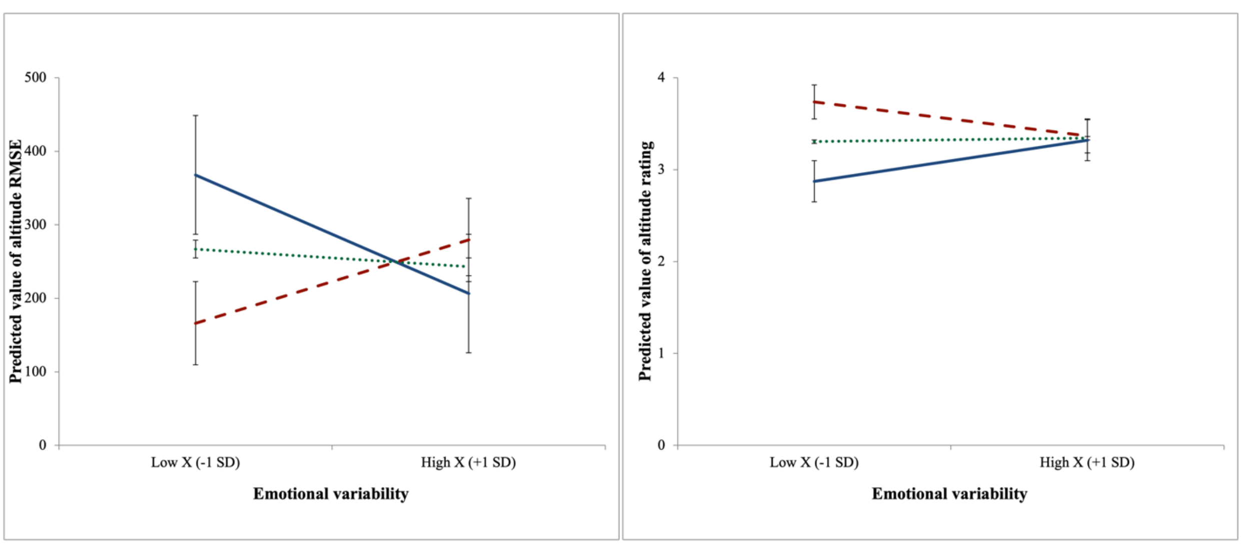
p=.01) and marginally for variance in altitude rating (R^2=-.699, F(1, 18)=-1.90, p=.07). Johnson-Neyman analysis revealed that association between emotional variability and altitude performance differed significantly from zero at -1SD of control ($p_{altitude\ RMSE}$=.004, $p_{altitude\ rating}$=.05). Simple slope analyses (Figure 22) indicated that participants who perceived less control and displayed lower emotional variability had less accurate performance. When participants perceived low control and exhibited high emotional variability, they achieved more accurate altitude performance. Descriptively, with lower emotional variability, perceived control has a greater effect on flying performance.

Figure 22

Moderating Effect of Perceived Control Between Emotional Variability and Altitude Performance During Session A

Low Control (-1 SD) Average Control High Control (+1 SD)

A) RMSE. B) Expert rating



Note. Error bars correspond to standard errors of performance metrics.

Session B

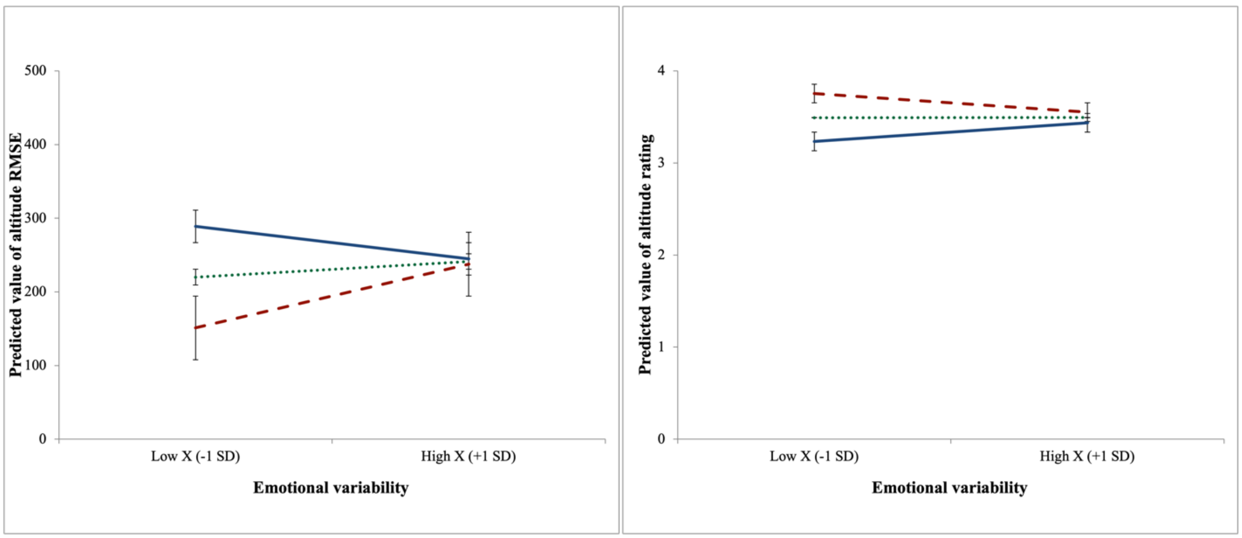
In session B, perceived control marginally and independently contributed to explaining variance in altitude RMSE (R^2=-.336, t=-2.02, p=.06). The interaction term emotional variability by perceived control had an antagonist interaction that marginally accounted for additional variation in altitude RMSE, R^2=.147, $F(1, 18)$=3.66, p=.07. For the antagonist interaction we refer to the association between the variables, where X and W independently associate in the same direction with Y, however, the interaction term XW have the opposite direction with Y (Hayes, 2022). In this case, emotional variability and perceived control independently had a negative association with RMSE, however, the interaction term between both shifted to positively associate with RMSE. Johnson-Neyman analysis indicated that the region where the association between emotional variability and altitude RMSE was marginally significantly different from zero at -1SD of control (p=.08). Simple slope analyses (Figure 23) show that participants who perceived less control and exhibited lower emotional entropy had less accurate altitude performance. When participants perceived low control and displayed high emotional variability, they had a more accurate altitude performance. Like the previous phases, with lower emotional variability, perceived control has a greater effect on flying performance.

Figure 23

Moderating Effect of Perceived Control Between Emotional Variability and Altitude Performance During Session B

Low Control (-1 SD) Average Control High Control (+1 SD)

A) RMSE. B) Expert rating



Note. Error bars correspond to standard errors of performance metrics.

Discussion

The findings of this study show that pilot trainees' perceived control and value over the task moderate the relationship between emotional variability and performance in a simulated flying task. These interactions varied across different training phases. Results revealed that the interaction between perceived value and emotional variability influenced altitude performance during the introduction phase. Surprisingly, participants who perceived high value over the task had a fair altitude performance despite level of emotional variability. Participants perceiving low value and displayed low emotional variability performed poorly, while those with high emotional variability performed more accurately. Results showed an unpredicted third profile, medium value, showing a significant interaction with emotional variability, like that of low value. Future studies can explore intermediate levels of emotional variability, following the Yerkes-Dodson law (1908), to better understand ideal levels of emotional variability in relation to performance accuracy.

Emotional variability appears to act as a stimulus for participants with low to medium value appraisals. More emotional variability correlated with more accurate flying performance in these cases. This finding aligns with the affect dynamics model, proposing that emotional disengagement relates to poorer performance when the task lacks stimulation to trigger emotional reactions (D'Mello & Graesser, 2012). Participants started the simulation with low perceived value in relation to their personal objectives, and the task was not stimulating enough to trigger a variety of emotions (i.e., low emotional variability), leading to poorer performance (D'Mello & Graesser, 2012; Pekrun & Perry, 2014).

Regarding altitude performance during session A and B, low perceived control in conjunction with low emotional variability related to poor performance, while low control with high emotional variability led to greater accuracy. This pattern supports the notion that low control with high emotional variability towards task culmination can yield better performance, sustaining emotional engagement (D'Mello & Graesser, 2012). Simple slopes showed that participants with high emotional variability performed relatively well, independently of the level of perceived control. The best performance was granted with high control and low emotional variability, and the worse performance was related to low control and low emotional variability.

Previous psychology research showed that uncontrolled fluctuating emotions and low perception of ownership over the task might negatively affect performance (Bailen et al., 2019; Li et al., 2021a, 2021b; Xu et al., 2016). However, our results show that emotional variability served as an adaptive reaction for participants with low perceived value and control over the task. Emotional variability can be paired with having more psychological flexibility (Kashdan & Rottenberg, 2010). Trainees likely adapted their emotional reactions according to the situational demands, showing functional self-regulation skills, resulting in more accurate performance (Kashdan & Rottenberg, 2010).

An interesting observation was that only one appraisal was significant in each training phase, with perceived value being significant for introduction phase, and perceived control being significant for session A and marginally for session B. Perceived control and value were significantly lower during introduction compared to sessions A and B. Additionally, previous analyses demonstrated that flying performance became more accurate as training phases advanced. Thus, the lack of significant interactions towards the end of the task might be explained by the improvement in flying performance and increase perceived control and value: with more practice participants were more accurate, had higher agency, and identified more alignment between their goals and the task (Pekrun, 2019).

This study poses limitations that should be accounted for in future studies. For instance, no statistically significant results were observed for heading performance. This might be due the manner that heading is manipulated. Controlling the yoke is like turning a car steering wheel to change directions, and ab-initio pilot trainees might be more familiar with this movement compared to pulling and pushing a yoke to control for altitude. Thus, heading performance was sustained, whereas the new skill of controlling altitude implied more learning. Moreover, the lack of significant findings might result from the small sample size, and consequent low effect size. The pool of volunteers meets the criteria to apply to aviation school; however, we recognize that the diversity in participants' previous experience likely influenced our results, limiting our generalizability. Our results confirmed that participants' background did not affect the variables in the moderation models with significant results. However, we found that participants with a master's degree reported less control during the introduction phase compared to participants with a bachelor's degree. This pattern suggests that, although results show an interesting direction, the experiment should be reproduced with pilot trainees enrolled in aviation school.

Conclusion

This study reveals that on-task emotional variability significantly influences flying performance of ab-initio trainees. Particularly, this relationship is explained when there are changes in perceived control and value over the task. This study contributes to understanding the role of emotional changes above and beyond traditionally studied emotional features, like stress and anxiety (Eysenck et al., 2007; Hart, 2006; Vine et al., 2015; Zheng et al., 2023a).

This study contributes to understanding the control-value theory of achievement emotions. Pekrun's theory (2019) typically explores control and value appraisals and performance in relation to discrete emotions (joy, anger, etc.); however, this study highlights that emotions dynamics during the learning activity interact with task success or failure (D'Mello & Graesser, 2012; Zheng et al., 2023a). Perceived value is recognized to relate to discrete emotions on professional simulated tasks; our results add that its interaction with emotional variability (gauging negative and positive emotions) can predict performance accuracy (Duffy et al., 2018).

Compared to previous studies, our results demonstrate that emotional variability might be adaptive when trainees have low perceived control or value (S. Li et al., 2021a; S. Li et al., 2021b; Xu et al., 2016). More emotional variability might reflect that trainees are investing efforts to adapt their emotions to task demands (Kashdan & Rottenberg, 2010). Understanding trainees' perceived control and value might be critical in early training stages when simulations are more relevant. Simulations are remarkably beneficial for junior trainees to familiarize themselves with their emotional reactions, deliberatively practice to improve performance, and recognize the importance of the task (Lajoie, 2021).

These conclusions can guide the use of simulations for professional training. Flight instructors can screen ab-initio pilot trainees perceived control and value over flying tasks. By knowing participants control and value, instructors can predict emotional patterns, and make personalized interventions to improve performance accuracy. To enhance perceived value, flight instructors are encouraged to understand trainees' motivations and create activities that align with trainees' objectives, and emphasize connections between tasks and their goals (Artino et al., 2012). To boost perceived control, instructors can offer choices between instructional activities with respect to perceived level of challenge that scaffold learners to achieve attainable goals, and provide timely constructive feedback (Artino et al., 2012). Flight training curriculum could incorporate emotion awareness and regulation techniques for trainees to familiarize themselves with emotions to further understand the impact on performance accuracy (Bailen et al., 2019; Gross, 2015).

References

Allsop, J., & Gray, R. (2014). Flying under pressure: Effects of anxiety on attention and gaze behavior in aviation. *Journal of Applied Research in Memory and Cognition*, *3*(2), 63–71. https://doi.org/10.1016/j.jarmac.2014.04.010

Artino, A. R., Holmboe, E. S., & Durning, S. J. (2012). Control-value theory: Using achievement emotions to improve understanding of motivation, learning, and performance in medical education: AMEE Guide No. 64. *Medical Teacher*, *34*(3), e148–e160. https://doi.org/10.3109/0142159X.2012.651515

Bailen, N. H., Green, L. M., & Thompson, R. J. (2019). Understanding emotion in adolescents: A review of emotional frequency, intensity, instability, and clarity. *Emotion Review*, *11*(1), 63–73. https://doi.org/10.1177/1754073918768878

Carden, S. W., Holtzman, N. S., & Strube, M. J. (2017). CAHOST: An excel workbook for facilitating the Johnson-Neyman technique for two-way interactions in multiple regression. *Frontiers in Psychology*, *8*, 1293. https://doi.org/10.3389/fpsyg.2017.01293

Cootes, T. F., & Taylor, C. J. (2001). Statistical models of appearance for medical image analysis and computer vision. *Medical Imaging 2001: Image Processing*, *4322*, 236–248. https://doi.org/10.1117/12.431093

D’Mello, S., & Graesser, A. (2012). Dynamics of affective states during complex learning. *Learning and Instruction*, *22*(2), 145–157. https://doi.org/10.1016/j.learninstruc.2011.10.001

Duffy, M. C., Lajoie, S., & Lachapelle, K. (2016). Measuring emotions in medical education: Methodological and technological advances within authentic medical learning environments. In S. Bridges, L. K. Chan, & C. E. Hmelo-Silver (Eds.), *Educational*

Technologies in Medical and Health Sciences Education (pp. 181–213). Springer. https://doi.org/10.1007/978-3-319-08275-2_10

Duffy, M. C., Lajoie, S. P., Pekrun, R., & Lachapelle, K. (2018). Emotions in medical education: Examining the validity of the Medical Emotion Scale (MES) across authentic medical learning environments. *Learning and Instruction*, 101150. https://doi.org/10.1016/j.learninstruc.2018.07.001

Ekman, P., & Rosenberg, E. L. (Eds.). (2005). *What the face reveals: Basic and applied studies of spontaneous expression using the facial action coding system (FACS)* (2nd ed.). Oxford University Press. 10.1093/acprof:oso/9780195179644.001.0001

Eysenck, M. W., Derakshan, N., Santos, R., & Calvo, M. G. (2007). Anxiety and cognitive performance: Attentional control theory. *Emotion*, *7*(2), 336–353. https://doi.org/10.1037/1528-3542.7.2.336

FAA Department of Transportation. (2003, September 17). *14 CFR 91.1059 -- Flight time limitations and rest requirements: One or two pilot crews.* Electronic Code of Federal Regulations. https://www.ecfr.gov/current/title-14/chapter-I/subchapter-F/part-91/subpart-K/subject-group-ECFRc17623c0e0be17e/section-91.1059

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*(2), 175–191. https://doi.org/10.3758/bf03193146

Gaetan, S., Dousset, E., Marqueste, T., Bringoux, L., Bourdin, C., Vercher, J.-L., & Besson, P. (2015). Cognitive workload and psychophysiological parameters during multitask activity in helicopter pilots. *Aerospace Medicine and Human Performance*, *86*(12), 1052–1057. https://doi.org/10.3357/AMHP.4228.2015

Gao, Z., & Xiang, P. (2008). College students' motivation toward weight training: An application of expectancy-value model. *Journal of Teaching in Physical Education*, *27*(3), 399–415. https://doi.org/10.1123/jtpe.27.3.399

Gray, R., Gaska, J., & Winterbottom, M. (2016). Relationship between sustained, orientated, divided, and selective attention and simulated aviation performance: Training & pressure effects. *Journal of Applied Research in Memory and Cognition*, *5*(1), 34–42. https://doi.org/10.1016/j.jarmac.2015.11.005

Gross, J. J. (2015). Emotion regulation: Current status and future prospects. *Psychological Inquiry*, *26*(1), 1–26. https://doi.org/10.1080/1047840X.2014.940781

Gruber, J., Kogan, A., Quoidbach, J., & Mauss, I. B. (2013). Happiness is best kept stable: Positive emotion variability is associated with poorer psychological health. *Emotion*, *13*(1), 1–6. https://doi.org/10.1037/a0030262

Hamman, W. R. (2004). Will simulation fly in medicine as it has in aviation? *Quality & Safety in Health Care*, *13*(5), 397–399. https://doi.org/10.1136/qhc.13.5.397

Harley, J. M., Bouchet, F., & Azevedo, R. (2012). Measuring learners' co-occurring emotional responses during their interaction with a pedagogical agent in MetaTutor. In S. A. Cerri, W. J. Clancey, G. Papadourakis, & K. Panourgia (Eds.), *Intelligent Tutoring Systems* (pp. 40–45). Springer. https://doi.org/10.1007/978-3-642-30950-2_5

Hart, S. G. (2006). Nasa-Task Load Index (NASA-TLX): 20 years later. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *50*(9), 904–908. https://doi.org/10.1177/154193120605000909

Hart, S. G., & Bortolussi, M. R. (1984). Pilot errors as a source of workload. *Special Issue: Aviation Psychology*, *26*(5), 545–556.

Hayes, A. F. (2022). Moderation analysis. In *Introduction to Mediation, Moderation, and Conditional Process Analysis: A Regression-Based Approach* (pp. 231–281). Guilford Press.

Herrera-Aliaga, E., & Estrada, L. D. (2022). Trends and innovations of simulation for twenty first century medical education. *Frontiers in Public Health*, *10*, 619769. https://doi.org/10.3389/fpubh.2022.619769

Hou, C., Zhu, G., & Yang, Y. (2023, June 10). *Beyond the cognitive dimension: Emotion patterns in productive and improvable knowledge building discourse*. Annual Meeting of the International Society of the Learning Sciences. https://doi.org/10.22318/cscl2023.333889

Houben, M., Van Den Noortgate, W., & Kuppens, P. (2015). The relation between short-term emotion dynamics and psychological well-being: A meta-analysis. *Psychological Bulletin*, *141*(4), 901–930. https://doi.org/10.1037/a0038822

Jack, R. E., Garrod, O. G. B., & Schyns, P. G. (2014). Dynamic facial expressions of emotion transmit an evolving hierarchy of signals over time. *Current Biology: CB*, *24*(2), 187–192. https://doi.org/10.1016/j.cub.2013.11.064

Jennings, S., Law, A., Bourgon, A., & Grenier, H. (2024, January). *Novel methodology for comparing student pilot flight performance with instructor pilot expert ratings*. AIAA SCITECH 2024 Forum, Orlando, Florida.

Karaca, Y., & Moonis, M. (2022). Shannon entropy-based complexity quantification of nonlinear stochastic process: Diagnostic and predictive spatiotemporal uncertainty of multiple sclerosis subgroups. In Y. Karaca, D. Baleanu, Y.-D. Zhang, O. Gervasi, & M. Moonis (Eds.), *Multi-Chaos, Fractal and Multi-Fractional Artificial Intelligence of Different*

Complex Systems (pp. 231–245). Academic Press. https://doi.org/10.1016/B978-0-323-90032-4.00018-3

Kashdan, T. B., & Rottenberg, J. (2010). Psychological flexibility as a fundamental aspect of health. *Clinical Psychology Review*, *30*(7), 865–878. https://doi.org/10.1016/j.cpr.2010.03.001

Kuppens, P. (2015). It's about time: A special section on affect dynamics. *Emotion Review*, *7*(4), 297–300. https://doi.org/10.1177/1754073915590947

Lajoie, S. P. (2021). Multimedia learning with simulations. In L. Fiorella & R. E. Mayer (Eds.), *The Cambridge handbook of multimedia learning* (3rd ed., pp. 461–471). Cambridge University Press. https://doi.org/10.1017/9781108894333.048

Laminar Research. (2022). *X-Plane 11 Desktop manual*. https://www.x-plane.com/manuals/desktop/

Li, S., Zheng, J., & Lajoie, S. P. (2021a). The frequency of emotions and emotion variability in self-regulated learning: What matters to task performance? *Frontline Learning Research*, *9*(4), 76–91. https://doi.org/10.14786/flr.v9i4.901

Li, S., Zheng, J., Lajoie, S. P., & Wiseman, J. (2021b). Examining the relationship between emotion variability, self-regulated learning, and task performance in an intelligent tutoring system. *Educational Technology Research and Development*, *69*(2), 673–692. https://doi.org/10.1007/s11423-021-09980-9

Li, T., & Lajoie, S. (2021). Predicting aviation training performance with multimodal affective inferences. *International Journal of Training and Development*, *25*(3), 301–315. https://doi.org/10.1111/ijtd.12232

Loijens, L., Krips, O., van Kuilenburg, H., den Uyl, M., & Ivan, P. (2015). *FaceReader version 6.1 Reference manual*. Noldus Information Technology.

Marques, E., Carim, G., Campbell, C., & Lohmann, G. (2023). Ab initio flight training: A Systematic literature review. *The International Journal of Aerospace Psychology*, *33*(2), 99–119. https://doi.org/10.1080/24721840.2022.2162405

Pekrun, R. (2019). Achievement emotions: A control-value theory perspective. In R. Patulny, A. Bellocchi, R. Olson, S. Khorana, J. McKenzie, & M. Peterie (Eds.), *Emotions in late modernity* (pp. 142-157). Routledge.

Pekrun, R., & Perry, R. P. (2014). Control-value theory of achievement emotions. In R. Pekrun & L. Linnenbrink-Garcia (Eds.), *International handbook of emotions in education* (pp. 120–141). Routledge.

Perry, R. P., Hladkyj, S., Pekrun, R. H., & Pelletier, S. T. (2001). Academic control and action control in the achievement of college students: A longitudinal field study. *Journal of Educational Psychology*, *93*, 776–789. https://doi.org/10.1037/0022-0663.93.4.776

Shannon, C. E. (1948). A mathematical theory of communication. *The Bell System Technical Journal*, *27*(3), 379–423. https://doi.org/10.1002/j.1538-7305.1948.tb01338.x

Transport Canada. (2019, November 4). *Standard 421—Flight Crew Permits, Licences and Ratings—Canadian Aviation Regulations (CARs)*. AARBH 15472583; AARBH. https://tc.canada.ca/en/corporate-services/acts-regulations/list-regulations/canadian-aviation-regulations-sor-96-433/standards/standard-421-flight-crew-permits-licences-ratings-canadian-aviation-regulations-cars

Vine, S. J., Uiga, L., Lavric, A., Moore, L. J., Tsaneva-Atanasova, K., & Wilson, M. R. (2015). Individual reactions to stress predict performance during a critical aviation incident. *Anxiety, Stress & Coping: An International Journal*, *28*(4), 467–477.

Viola, P., & Jones, M. J. (2004). Robust real-time face detection. *International Journal of Computer Vision*, *57*(2), 137–154. https://doi.org/10.1023/B:VISI.0000013087.49260.fb

Wigfield, A., & Eccles, J. S. (2000). Expectancy–value theory of achievement motivation. *Contemporary Educational Psychology*, *25*(1), 68–81. https://doi.org/10.1006/ceps.1999.1015

Xu, S., Martinez, L. R., Van Hoof, H., Eljuri, M. I., & Arciniegas, L. (2016). Fluctuating emotions: Relating emotional variability and job satisfaction. *Journal of Applied Social Psychology*, *46*(11), 617–626. https://doi.org/10.1111/jasp.12390

Yerkes, R. M., & Dodson, J. D. (1908). The relation of strength of stimulus to rapidity of habit formation. *Journal of Comparative Neurology & Psychology*, *18*, 459–482. https://doi.org/10.1002/cne.920180503

Zafeiriou, S., Zhang, C., & Zhang, Z. (2015). A survey on face detection in the wild: Past, present and future. *Computer Vision and Image Understanding*, *138*, 1–24. https://doi.org/10.1016/j.cviu.2015.03.015

Zheng, J., Li, S., & Lajoie, S. P. (2023a). A review of measurements and techniques to study emotion dynamics in learning. In V. Kovanovic, R. Azevedo, D. C. Gibson, & D. lfenthaler (Eds.), *Unobtrusive observations of learning in digital environments* (pp. 7–29). Springer. https://doi.org/10.1007/978-3-031-30992-2_2

Appendix B

Control and Value Questionnaires Adapted for Flying Tasks

Adaptation of Academic Control Scale

1. I have a great deal of control over the flying task I just did.
2. I feel like the more effort I put it, the better I did at this flying task.
3. No matter how hard I tried, I could not have done this task better (R).

Adaptation of the Usefulness, Importance, and Interest Items from the Expectancy-Value Questionnaire

1. I think I can use what I learned from doing this flying task to other things in other situations.
2. This task is useful.
3. I had fun doing this task.
4. I find the task interesting.
5. It is important to me to learn this flying task.
6. For me, being good in this task is important.

Note. R stands for reverse scoring.

Appendix C

Differences in Control and Value According to Participants' Background

Training Phases – Educational Levels

Table C-1

Perceived Control and Value Differences According to Educational Background

Performance measure	Practice phase		Mean	SD	X^2	p
Control	Introduction	High school	3.77	0	10.92	.004
		Bachelors	3.97	0.33		
		Masters	3.45	0.45		
	Session A	High school	3.66	0.47	4.86	.088
		Bachelors	4.29	0.45		
		Masters	3.77	0.62		
	Session B	High school	4	0.47	1.112	.574
		Bachelors	4.22	0.78		
		Masters	4.02	0.54		
Value	Introduction	High school	3.69	0.90	.991	.609
		Bachelors	4.01	0.43		
		Masters	3.78	0.30		
	Session A	High school	3.75	0.58	1.294	.524
		Bachelors	4.11	0.51		
		Masters	3.94	0.62		
	Session B	High school	4.16	1.17	2.275	.321
		Bachelors	4.24	0.52		
		Masters	3.86	0.31		

Note. Results based on Kruskal-Wallis tests statistic. SD=Standard Deviation.

Table C-2

Differences in Emotional Variability according to Educational Background

	Training phases		Mean	SD	KW	*p*
Emotional variability	Introduction	High school	1.89	.05	1.817	.403
		Bachelors	1.71	.25		
		Masters	1.44	.71		
	Session A	High school	1.78	.11	1.227	.541
		Bachelors	1.54	.42		
		Masters	1.62	.57		
	Session B	High school	1.24	.36	4.522	.104
		Bachelors	1.53	.35		
		Masters	1.68	.59		

Note. Results based on Kruskal-Wallis tests statistic. SD=Standard Deviation.

Training Phases – Experience with Simulators

Table C-3

Differences in Performance according to Experience with Simulators

Performance measure	Training phase	Experience with simulators	Mean	SD	U	*p*
Control	Introduction	Yes	3.95	0.69	54	.401
		No	3.58	0.34		
	Session A	Yes	4.22	0.83	72	.155
		No	3.88	0.48		
	Session B	Yes	4.11	0.58	50	.973
		No	4.09	0.65		
Value	Introduction	Yes	3.67	0.42	28	.283
		No	3.91	0.39		
	Session A	Yes	3.77	0.51	33.50	.227
		No	4	0.40		
	Session B	Yes	3.88	0.44	38.50	.392
		No	4.08	0.52		

Note. Results based on Mann-Whitney U tests statistic. SD=Standard Deviation.

Table C-4

Emotional Variability Differences according to Experience with Simulators

	Difficulty level	Experience with simulators	Mean	SD	U	*p*
Emotional variability	Low	Yes	1.84	0.32	59.50	.189
		No	1.63	0.30		
	Medium	Yes	1.59	0.37	43	1.0
		No	1.51	0.55		
	High	Yes	1.68	0.45	42	1.0
		No	1.68	0.34		

Note. Results based on Mann-Whitney U tests statistic. SD=Standard Deviation.

Chapter 5. Final Discussion

This dissertation started by illustrating the existing gap between pilots' supply and flight demand, which is expected to continue to grow if the aviation industry does not implement any changes (Murray et al., 2022). A proposed solution to fill-in this gap is to improve training methods for educating pilots who can perform accurately, even in unexpected events.

When exploring individual differences that contribute to performance accuracy, aviation research has focused mainly on cognitive factors, such as workload and attentional control (Eysenck et al., 2007; Hart, 2006). On the other hand, educational theories are increasingly describing how affective processes have a significant role in performance accuracy (D'Mello & Graesser, 2012; Pekrun, 2019; Zheng et al., 2023a). Current approaches in aviation mainly account for negative-activating affective states, like stress and anxiety, as a subordinate factor to explain cognitive processing, and their resultant performance accuracy (Eysenck et al., 2007; Hart, 2006). However, studies in professional training are showing that different types of emotions, including positive and negative, activating and deactivating, have specific effects on performance accuracy and decision making (Artino et al., 2012). Yet the effect of these emotions will be specific for the domain and task characteristics (Duffy et al., 2018).

The objective of this dissertation is to provide theoretically and empirically sound research to inform new training approaches for growing the pool of pilots, attempting to meet the current flights demand. The findings of this dissertation successfully connected aviation and educational approaches to map contextual qualities (i.e., performance measures in the context of flight simulations) with learners' individual characteristics, such as emotional responses, to illustrate the relationships between the dynamic nature of emotional states and how emotions influence flight performance in simulations. The findings from this dissertation can inform the creation of innovative pilot training research and curriculum development that take emotional experience and subjective appraisals of the task into consideration to enhance performance accuracy. The following sections summarize the contributions of this dissertation, considering theoretical, methodological, and practical inputs, as well as limitations and future directions.

Theoretical Contributions

This dissertation makes important contributions to the theoretical understanding of emotion dynamics in performance accuracy in the context of training pilots with flight simulations. In Chapter 2, we synthesized studies exploring the relationship between affective responses and flying performance when using flight simulations. This synthesis demonstrated a significant growth in research exploring the impact of affect in flight training since the beginning of the 21st century. The synthesis in Chapter 2 demonstrates that, due to the complexity of flying, intense and not well managed emotions can be distractions, leading to human error. Most of the research in this domain attempts to understand affect in unexpected and intense events, to improve pilots' decision making. For instance, there is a considerable interest in using continuous and non-intrusive measures of affect and performance to allow pilots to fully concentrate on the task. Moreover, the results show that unmanaged, and intense anxiety, stress, and surprise tend to be detrimental for flying performance; however, there are instances in which controllable levels of these affective reactions might activate the pilots to solve the task. In this regard, the synthesis in Chapter 2 shows that training pilots to use performance and affect management techniques is beneficial for having more accurate performance. We emphasize that these results are contextually grounded. Thus, the task characteristics, like the type of simulations and social interactions with peer pilots and instructors, also have an effect in the affective experience.

The review in Chapter 2 emphasizes that affect and performance are continuous and co-occurring processes. This finding was used as a guideline to create the empirical studies in Chapter 3 and 4, that assess emotions and performance in a continuous, and non-invasive manner. Moreover, the findings in Chapter 2 show that there are gaps for understanding the degree of functionality of affective reactions in pilot performance. Namely, at what degree can positive and negative emotions activate pilots to solve the task, or rather distract them from meeting the task demands.

Therefore, Chapter 3 and 4 contribute to the understanding of the role that emotions play in flight training. This dissertation contributes to understanding the connection between emotion dynamics, like intensity and variability, in flying performance accuracy. Moreover, due to the growth in technology in aviation, a main contribution of this dissertation is the study of beginner pilot trainees as they use technology-based flight simulations to guide instructional implications for training accurate pilots (Salas et al., 2010).

According to the literature analyzed in Chapter 2, empirical studies have explored emotional and performance changes across time or according to difficulty levels, but not together. Chapter 3 contributes to the literature by showing a detailed analysis of performance and emotional changes across training phases and difficulty levels. Results show that beginner participants became more accurate as training phases advanced, and they were less accurate during high-difficulty tasks. However, the main contribution of this study was identifying the emotional changes that participants had during the flight simulations.

Chapter 3 contributes to understanding the dynamic features of emotions in professional training using a multimodal approach. To our knowledge, this is the first study that explores three dynamic features of emotions for professional training, namely, emotional frequency, intensity (inferred from intense changes in electrodermal activity and facial expression of emotions), and variability (inferred from fluctuations in intense facial expression of multiple emotions). Chapter 3 provides a detailed description of the emotional experience of beginner pilot trainees across training phases and difficulty levels in a flight simulation.

Compared to previous studies exploring affect in flight training, the results of this study pose different patterns in frequency of discrete emotions during flight training. Particularly, we mentioned above that models exploring affect in flight training focus on cognitive processes, and negative and activating affect, like stress, anxiety, and surprise. However, our results suggest that participants more frequently and persistently (intensely) express neutrality and anger, interpreted as deep focus; and participants expressed happiness, fear (paired with anxiety) and disgust less frequently. Thus, our findings demonstrate that a wider range of affective responses might have a function in flight training and performance accuracy. However, it should be noted that surprise had an intermediate frequency across training phases and difficulty levels. Aligned with previous studies, we believe that surprise should be researched in more detail as its immediate effect in performance might be negative, by distracting participants, yet if the impasse is solved successfully, the increase in physiological arousal might stimulate learners to solve the unexpected change (D'Mello & Greasser, 2012; Landman et al., 2020).

Findings of Chapter 3 also demonstrate that dynamic features of emotions provide rich information about the impact of emotions in performance accuracy. Emotions in learning are traditionally studied as the frequency of discrete emotions (S. Li et al., 2021a). An added benefit of analyzing emotions dynamically is that emotions can be interpreted more globally, emphasizing their functionality, independently of their valence, attempting to avoid the misunderstanding that positive emotions are "better" than negative ones.

We recognize that Chapter 3 had some limitations. First, the relationship between flying performance and emotional variability was descriptively discussed, but the correlation between both was not explored. Moreover, despite using a multimodal approach, the study presented in Chapter 3 did not account for participants' subjective experience. Thus, using Pekrun's CVT (2019), Chapter 4 was designed to explore the relationship between emotional variability and flying performance, by exploring if this relationship varied when participants reported different levels of control and value over the task.

The results of Chapter 4 demonstrate that beginner pilot trainees' emotional variability and flying performance are correlated when moderated by perceived control and value over the task. When participants reported a high perceived control or value, their performance was generally good regardless of experiencing low or high emotional variability. However, emotional variability had an adaptive function for participants who reported low perceived value or control. When participants had more emotional variability and had low perceived control or value, they had more accurate performance than those with low emotional variability. These findings have multiple contributions for psychological, educational, and aviation research.

For psychology, emotional variability has been mostly studied as a psychopathological symptom (Bailen et al., 2019; Thompson et al., 2012); the few studies that have explored emotional variability in neurotypical contexts have concluded that it tends to be correlated to unfavourable outcomes (S. Li et al, 2021a; S. Li et al., 2021b; Xu et al., 2016). However, the results in Chapter 4 demonstrate that emotional variability likely has adaptive functions (Kashdan & Rottenberg, 2010; Xu et al., 2016). The fluctuations in emotional variability might indicate a constant effort to adapt to the situation, thus being perceived as a signal of self-regulation efforts (Kashdan & Rottenberg, 2010; Kuppens, 2015). In terms of education, results of Chapter 4 contribute to understanding the interaction between learners' emotions and perceived control and value to explain their performance outcomes (Duffy et al., 2020; Pekrun, 2019). Low perceived control and value and high emotional variability, independently, tend to be associated to poorer performance (Li et al., 2021a, 2021b; Pekrun & Perry, 2014). Yet our results demonstrate that high emotional variability might be especially beneficial for learners with lower control and value over the task (Kashdan & Rottenberg, 2010; Xu et al., 2016). Finally, the results of Chapter 4 demonstrate that the subjective perceptions of pilot trainees interact with their affective responses, explaining a portion of their performance accuracy (Hart & Bortolussi, 1984; Vine et al., 2015). In that regard, the methods used in this dissertation show contributions to methodologies that can be adapted for psychological, educational, professional training, and aviation research.

Methodological Importance

Beyond the theoretical contributions, this dissertation also contributes to how specific multimodal measures can be used for assessing performance accuracy and emotion dynamics in the context of simulated flying tasks (Harley, 2016; T. Li & Lajoie, 2021; Zheng et al., 2023a). It should be recognized that the measurements of flying performance were mainly created by the co-authors of Chapter 3 and 4, Dr. Law, Jennings, and subject-matter expert Bourgon (see Jennings et al., 2024). As found in the studies synthesized in Chapter 2, flight performance tends to be studied by an objective or a subjective measure in isolation (i.e., Allsop & Gray, 2014; McClernon & Miller, 2011; Vallès-Català et al., 2021; Wang et al., 2016). Therefore, a methodological contribution of this dissertation is that it accounted for both objective (i.e., root-mean-square-error used to calculate accuracy based on log file data within the simulated flying tasks) and subjective (experienced instructor ratings) measures to understand performance accuracy (Jennings et al., 2024).

The main methodological contribution of this dissertation is showing that dynamic features of emotions can be analyzed in a multimodal, quantifiable, and non-invasive manner (Harley, 2016; Zheng et al., 2023a). Dynamic features were theoretically defined, allowing us to conceptualize and distinguish different qualities of emotions, including frequency, intensity i.e., strength of a single variable like skin conductance responses, and variability, implying fluctuations among multiple emotions (Bailen et al., 2019; Zheng et al., 2023a). Past research in this area only examined the static nature of emotion by using frequency of an emotional reaction as an indicator, counting the occurrence of an emotional reaction in a designated period, using two measures: facial expression of basic emotions and skin conductance responses (i.e., T. Li & Lajoie, 2021). These two measurements were only accounted for when they had a high intensity, as being significantly more persistent than other emotions in the case of facial expression and noting a significant change in electrodermal activity (Boucsein et al., 2012; Harley et al., 2019b; Loijens et al., 2015). Moreover, facial expression of multiple dominant emotions was used to understand how pilot trainees' emotions changed across the tasks (S. Li et al, 2021a; S. Li et al., 2021b).

A critique in Chapter 2 was that some studies exploring affect in flight training lack theoretical guidance to select their methods, and thus limit the interpretability of results. In that regard, this dissertation demonstrates theory-grounded methods to understand the emotional experience of pilot trainees. This dissertation followed practices used in the aviation domain, emphasizing the use of non-invasive and continuous measures to assess both affect and performance (Drinkwater et al., 1968; Gaetan et al., 2015; T. Li & Lajoie, 2021; Rosa et al., 2021, 2022; Silva et al., 2009; Tichon et al., 2014). Particularly, this dissertation demonstrates the significance of using multimodal measures that combine objective and subjective assessment of pilot trainees experience, leading to a rich understanding of the interaction between affect and flying performance (T. Li & Lajoie, 2021). Lastly, the manuscripts are the first studies that quantitatively examines emotional variability in the context of flight training (S. Li et al, 2021a; S. Li et al., 2021b; Xu et al., 2016).

Practical Contributions

The results of this dissertation have implications that can contribute to the field of pilot training and have the potential to generalize to other high-stakes professions (Azher et al., 2023; Hamman, 2004). The results demonstrate that ab-initio pilot trainees, with little to no experience using flight simulations, significantly improved their performance by practicing in a flight simulator, showing the benefits of training in a safe and authentic environment (Jorna, 1993; Lajoie, 2021). Specifically, the results show that pilot trainees' emotional experience and subjective appraisals of the task influenced their performance accuracy (Duffy et al., 2020; Pekrun, 2019). This dissertation invites pilot training instructors and curricular creators to account for the emotional experience of learners. Instructors can align activities with learners' objectives, offering a range of activities, and providing individualized and constructive feedback (Artino & Jones, 2012). Moreover, pilot training curricula can include instruction of emotion awareness and emotional regulation techniques (Bailen et al., 2019).

One of the main arguments in this dissertation is that all affective reactions have a function. However, expression of emotions is discouraged in high-stakes and competitive professional domains, like aviation and health sciences (Duffy et al., 2016). The results of this dissertation demonstrate that, to judge the functionality of an emotional reaction, it is key to understand the interaction between type of emotion, intensity, and variability, as well as considering environmental and task characteristics (Damasio, 2005; Pekrun, 2019). The results of this dissertation demonstrate that behavioral expression of emotion can signal key moments to create interventions to improve performance (Calvo & D'Mello, 2010; Harley et al., 2017). For instance, our results suggest that (behaviorally) expressing a wider range of emotions might be beneficial for learners with low perceived control and value over the task (Duffy et al., 2020; Kashdan & Rottenberg, 2010; S. Li et al., 2023a; S. Li et al., 2023b). Thus, these results can invite instructors in high-stakes domains to create environments for learners to express their emotions in a safe manner, attempting to improve their formation as professionals (Artino & Jones, 2012; Skibniewski et al., 2015).

Limitations and Future Directions

The manuscripts presented in this dissertation have limitations that require attention. However, these limitations can be used to inform future research. This dissertation contributed to earlier research themes found in chapter 2 that study the relationship between affect and flying performance, by using continuous and non-invasive measures of emotions, examining cognitive appraisals, exploring changes in anxiety (paired with facial expression of fear), and surprise. However, this dissertation did not explore the role of social interactions, nor affect regulation techniques (Lobczowski, 2022). Our results suggest that the presence or absence of feedback might have a significant role for explaining beginner pilot trainees emotional experience; therefore, future research can explore the implications of the type of feedback, and presence of an instructor observing the task (Krahenbuhl et al., 1981; Skibniewski et al., 2015). Additionally, future research could investigate the impact of teaching emotion regulation techniques to pilot trainees (Harley et al., 2019a; Landman et al., 2020; McClernon et al., 2011).

A limitation illustrated in Chapters 3 and 4 is the small sample size, hampering the generalizability of the interpretations of the results, and limiting the generalization of result to student pilots (Cohen, 1988). Participants had diverse backgrounds, which discreetly influenced the patterns observed in participants. For instance, participants with a bachelors' degree expressed more anger, less neutrality, and higher control than participants with a master's degree. However, we confirmed that previous experience with flight simulators did not change the patterns in performance, nor emotional responses, despite potential ease using the simulator in the current experiment (Althubaiti, 2016). Although our objective was to mimic beginner pilot trainees, it is likely that student pilots have more experience and interest in flight simulations than our participant pool, which might result in differences in performance accuracy and higher control and value ratings (Pekrun, 2019; Vine et al., 2015). More specifically, the circumstances under which perceived control and value explain the relationship between emotional variability and performance accuracy might differ for student pilots already enrolled in aviation school, despite having little previous experience (Duffy et al., 2020). Therefore, a future direction would be to reproduce our research with student pilots and a large sample size to understand if the results replicate and whether changes in prior knowledge, control and value may vary according to different participant pools (Duffy et al., 2020). Similarly, future research could model the reactions and regulation strategies of experienced pilots to create curricular adaptations based on an expert model (Ericsson, 2006; Lajoie et al., 2020).

Another limitation of this study is the lack of self-reported measures of emotions, thus lacking a grounded-truth for interpreting participants' experience (Harley, 2016). Although we recognize this limitation, it should be noted that we decided to take this risk due to the complexity of the InLook project, which involved additional physiological and behavioral measures (such as electroencephalograms and eye-tracking). Adding self-reports to the methodology would have required more time and cognitive effort for participants who were already involved in a long experiment (6h) (Dismukes, 2010). With these considerations, a future direction is to conduct a follow-up study that concentrates on the emotional experiences of beginner pilot trainees, that additionally includes self-reported measures of emotions like questionnaires and interviews, to better understand their emotion regulation techniques, and testing if these are correlated to emotional variability (S. Li et al., 2023a; S. Li et al., 2023b; T. Li & Lajoie, 2021).

Our results show that neutral and angry facial expressions are the more prevalent among pilot trainees (see Chapter 3). Based on previous research, we interpreted both emotions as a demonstration of focus, in which neutral shows that the learners' effort is fully on the task, or expressing a frowning face as a reflection of high concentration (Harley et al., 2012; T. Li & Lajoie, 2021). However, these continue to be assumptions as we lacked a self-report to confirm participants' subjective experience (Harley, 2016). Thus, a future direction of research is to do a bottom-up exploration using machine learning techniques for analyzing participants' facial expressions based on action units, grouping similar expressions, and labelling emotions a posteriori, contrary to the approach taken in this dissertation (D'Mello et al., 2010). Specifically, it might be relevant to use labels based on academic or epistemic emotions that are more pertinent for the context of professional training (Zheng et al., 2023b).

Concluding Remarks

The findings in this dissertation have compelling theoretical, methodological, and practical implications for research on emotion dynamics and performance accuracy when training with flight simulators. The findings indicate that dynamic features of emotions have adaptive implications for flight performance accuracy. Results demonstrate that some emotional reactions (such as being neutral) and experiencing more emotional variability with low perceived control or value over the task might be beneficial for having a more accurate performance. The results of this dissertation demonstrate the adaptive functionality of emotions in the context of flight training which can lead to future research and instruction that utilizes these findings to improve flight training. In this regard, the findings of this dissertation can guide instructional adaptations for training pilots to help them recognize and manage their emotions during flights, to enhance flight performance. Improving training opportunities can increase the quality and quantity of the pilots needed to meet the current world demands.

References

Allsop, J., & Gray, R. (2014). Flying under pressure: Effects of anxiety on attention and gaze behavior in aviation. *Journal of Applied Research in Memory and Cognition*, *3*(2), 63–71. https://doi.org/10.1016/j.jarmac.2014.04.010

Althubaiti, A. (2016). Information bias in health research: Definition, pitfalls, and adjustment methods. *Journal of Multidisciplinary Healthcare*, *9*, 211–217. https://doi.org/10.2147/JMDH.S104807

Artino, A. R., Holmboe, E. S., & Durning, S. J. (2012). Control-value theory: Using achievement emotions to improve understanding of motivation, learning, and performance in medical education: AMEE Guide No. 64. *Medical Teacher*, *34*(3), e148–e160. https://doi.org/10.3109/0142159X.2012.651515

Artino, A. R., & Jones, K. D. (2012). Exploring the complex relations between achievement emotions and self-regulated learning behaviors in online learning. *The Internet and Higher Education*, *15*(3), 170–175. https://doi.org/10.1016/j.iheduc.2012.01.006

Azher, S., Cervantes, A., Marchionni, C., Grewal, K., Marchand, H., & Harley, J. M. (2023). Virtual simulation in nursing education: Headset virtual reality and screen-based virtual simulation offer a comparable experience. *Clinical Simulation in Nursing*, *79*, 61–74. https://doi.org/10.1016/j.ecns.2023.02.009

Bailen, N. H., Green, L. M., & Thompson, R. J. (2019). Understanding emotion in adolescents: A review of emotional frequency, intensity, instability, and clarity. *Emotion Review*, *11*(1), 63–73. https://doi.org/10.1177/1754073918768878

Boucsein, W., Fowles, D. C., Grimnes, S., Ben-Shakhar, G., Walton, T. Roth, Dawson, M. E., & Filion, D. L. (2012). Publication recommendations for electrodermal measurements. *Psychophysiology*, *49*(8), 1017–1034. https://doi.org/10.1111/j.1469-8986.2012.01384.x

Cohen, J. (1988). The Analysis of Variance. In *Statistical Power Analysis for the Behavioral Sciences* (2nd ed.). Routledge.

Damasio, A. R. (2005). *Descartes' error: Emotion, reason, and the human brain*. Penguin Books.

Dismukes, R. K. (2010). Understanding and analyzing human error in real-world operations. In E. Salas & D. Maurino (Eds.), *Human factors in aviation* (2nd ed., pp. 335–374). Academic Press. https://doi.org/10.1016/B978-0-12-374518-7.00011-0

D'Mello, S., & Graesser, A. (2012). Dynamics of affective states during complex learning. *Learning and Instruction*, *22*(2), 145–157. https://doi.org/10.1016/j.learninstruc.2011.10.001

D'Mello, S., Lehman, B., & Person, N. (2010). Monitoring affect states during effortful problem solving activities. *International Journal of Artificial Intelligence in Education*, *20*(4), 361–389. https://doi.org/10.3233/JAI-2010-012

Drinkwater, B. L., Flint, M. M., Clel, & Troy S. (1968). Somatic responses and performance levels during anticipatory physical-threat stress. *Perceptual and Motor Skills*, *27*(2), 539–552. https://doi.org/10.2466/pms.1968.27.2.539

Duffy, M. C., Lajoie, S., & Lachapelle, K. (2016). Measuring emotions in medical education: Methodological and technological advances within authentic medical learning environments. In S. Bridges, L. K. Chan, & C. E. Hmelo-Silver (Eds.), *Educational Technologies in Medical and Health Sciences Education* (pp. 181–213). Springer. https://doi.org/10.1007/978-3-319-08275-2_10

Duffy, M. C., Lajoie, S. P., Pekrun, R., & Lachapelle, K. (2018). Emotions in medical education: Examining the validity of the Medical Emotion Scale (MES) across authentic medical learning environments. *Learning and Instruction*, 101150. https://doi.org/10.1016/j.learninstruc.2018.07.001

Ericsson, K. A. (2006). The influence of experience and deliberate practice on the development of superior expert performance. In K. A. Ericsson, N. Charness, P. J. Feltovich, & R. R. Hoffman (Eds.), *The Cambridge handbook of expertise and expert performance* (pp. 683–704). Cambridge University Press. https://doi.org/10.1017/CBO9780511816796.038

Eysenck, M. W., Derakshan, N., Santos, R., & Calvo, M. G. (2007). Anxiety and cognitive performance: Attentional control theory. *Emotion*, *7*(2), 336–353. https://doi.org/10.1037/1528-3542.7.2.336

Gaetan, S., Dousset, E., Marqueste, T., Bringoux, L., Bourdin, C., Vercher, J.-L., & Besson, P. (2015). Cognitive workload and psychophysiological parameters during multitask activity in helicopter pilots. *Aerospace Medicine and Human Performance*, *86*(12), 1052–1057. https://doi.org/10.3357/AMHP.4228.2015

Hamman, W. R. (2004). Will simulation fly in medicine as it has in aviation? *Quality & Safety in Health Care*, *13*(5), 397–399. https://doi.org/10.1136/qhc.13.5.397

Harley, J. M. (2016). Measuring emotions: A survey of cutting edge methodologies used in computer-based learning environment research. In S. Y. Tettegah & M. Gartmeier (Eds.), *Emotions, technology, design, and learning* (pp. 89–114). Academic Press. https://doi.org/10.1016/B978-0-12-801856-9.00005-0

Harley, J. M., Bouchet, F., & Azevedo, R. (2012). Measuring learners' co-occurring emotional responses during their interaction with a pedagogical agent in MetaTutor. In S. A. Cerri,

W. J. Clancey, G. Papadourakis, & K. Panourgia (Eds.), *Intelligent tutoring systems* (pp. 40–45). Springer. https://doi.org/10.1007/978-3-642-30950-2_5

Harley, J. M., Lajoie, S. P., Frasson, C., & Hall, N. C. (2017). Developing Emotion-Aware, Advanced Learning Technologies: A Taxonomy of Approaches and Features. *International Journal of Artificial Intelligence in Education*, *27*(2), 268–297. https://doi.org/10.1007/s40593-016-0126-8

Hart, S. G., & Bortolussi, M. R. (1984). Pilot errors as a source of workload. *Special Issue: Aviation Psychology*, *26*(5), 545–556.

Hart, S. G. (2006). Nasa-Task Load Index (NASA-TLX); 20 Years Later. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *50*(9), 904–908. https://doi.org/10.1177/154193120605000909

Jennings, S., Law, A., Bourgon, A., & Grenier, H. (2024, January). *Novel methodology for comparing student pilot flight performance with instructor pilot expert ratings*. AIAA SCITECH 2024 Forum, Orlando, Florida.

Jorna, P. G. (1993). Heart rate and workload variations in actual and simulated flight. *Special Issue: Psychophysiological Measures in Transport Operations*, *36*(9), 1043–1054. https://doi.org/10.1080/00140139308967976

Kashdan, T. B., & Rottenberg, J. (2010). Psychological flexibility as a fundamental aspect of health. *Clinical Psychology Review*, *30*(7), 865–878. https://doi.org/10.1016/j.cpr.2010.03.001

Krahenbuhl, G. S., Marett, J. R., & Reid, G. B. (1978). Task-specific simulator pretraining and in-flight stress of student pilots. *Aviation, Space, and Environmental Medicine*, *49*, 1107–1110.

Kuppens, P. (2015). It's about time: A special section on affect dynamics. *Emotion Review*, *7*(4), 297–300. https://doi.org/10.1177/1754073915590947

Lajoie, S. P. (2021). Multimedia learning with simulations. In L. Fiorella & R. E. Mayer (Eds.), *The Cambridge handbook of multimedia learning* (3rd ed., pp. 461–471). Cambridge University Press. https://doi.org/10.1017/9781108894333.048

Lajoie, S. P., Pekrun, R., Azevedo, R., & Leighton, J. P. (2020). Understanding and measuring emotions in technology-rich learning environments. *Learning and Instruction*, *70*, 101272. https://doi.org/10.1016/j.learninstruc.2019.101272

Li, S., Zheng, J., & Lajoie, S. P. (2021a). The frequency of emotions and emotion variability in self-regulated learning: What matters to task performance? *Frontline Learning Research*, *9*(4), 76–91. https://doi.org/10.14786/flr.v9i4.901

Li, S., Zheng, J., Lajoie, S. P., & Wiseman, J. (2021b). Examining the relationship between emotion variability, self-regulated learning, and task performance in an intelligent tutoring system. *Educational Technology Research and Development*, *69*(2), 673–692. https://doi.org/10.1007/s11423-021-09980-9

Li, T., & Lajoie, S. (2021). Predicting aviation training performance with multimodal affective inferences. *International Journal of Training and Development*, *25*(3), 301–315. https://doi.org/10.1111/ijtd.12232

Lobczowski, N. G. (2022). Capturing the formation and regulation of emotions in collaborative learning: The FRECL coding procedure. *Frontiers in Psychology*, *13*, 846811. https://doi.org/10.3389/fpsyg.2022.846811

McClernon, C. K., & Miller, J. C. (2011). Variance as a measure of performance in an aviation context. *The International Journal of Aviation Psychology*, *21*(4), 397–412. https://doi.org/10.1080/10508414.2011.606765

Murray, G., Heilakka, R., & Medland, A. (2022, November 1). *The pilot shortage will only get worse as the decade progresses*. BRINK – Conversations and Insights on Global Business. https://www.brinknews.com/the-pilot-shortage-will-only-get-worse-as-the-decade-progresses/

Pekrun, R. (2019). Achievement emotions: A control-value theory perspective. In R. Patulny, A. Bellocchi, R. Olson, S. Khorana, J. McKenzie, & M. Peterie (Eds.), *Emotions in Late Modernity* (pp. 142-157). Routledge.

Pekrun, R., & Perry, R. P. (2014). Control-value theory of achievement emotions. In R. Pekrun & L. Linnenbrink-Garcia (Eds.), *International handbook of emotions in education* (pp. 120–141). Routledge.

Rosa, E., Gronkvist, M., Kolegard, R., Dahlstrom, N., Knez, I., Ljung, R., & Willander, J. (2021). Fatigue, emotion, and cognitive performance in simulated long-duration, single-piloted flight missions. *Aerospace Medicine and Human Performance*, *92*(9), 710–719. https://doi.org/10.3357/AMHP.5798.2021

Rosa, E., Lyskov, E., Gronkvist, M., Kolegard, R., Dahlstrom, N., Knez, I., Ljung, R., & Willander, J. (2022). Cognitive performance, fatigue, emotional, and physiological strains in simulated long-duration flight missions. *Military Psychology*, *34*(2), 224–236. https://doi.org/10.1080/08995605.2021.1989236

Silva, D. C., Vinhas, V., Reis, L. P., & Oliveira, E. (2009). Biometric emotion assessment and feedback in an immersive digital environment. *International Journal of Social Robotics*, *1*(4), 307–317. https://doi.org/10.1007/s12369-009-0029-z

Skibniewski, F. W., Dziuda, L., Baran, P. M., Krej, M. K., Guzowski, S., Piotrowski, M. A., & Truszczynski, O. E. (2015). Preliminary results of the LF/HF ratio as an indicator for estimating difficulty level of flight tasks. *Aerospace Medicine and Human Performance*, *86*(6), 518–523. https://doi.org/10.3357/AMHP.4087.2015

Thompson, R. J., Mata, J., Jaeggi, S. M., Buschkuehl, M., Jonides, J., & Gotlib, I. H. (2012). The everyday emotional experience of adults with major depressive disorder: Examining emotional instability, inertia, and reactivity. *Journal of Abnormal Psychology*, *121*(4), 819–829. https://doi.org/10.1037/a0027978

Tichon, J. G., Wallis, G., Riek, S., & Mavin, T. (2014). Physiological measurement of anxiety to evaluate performance in simulation training. *Cognition, Technology & Work*, *16*(2), 203–210.

Vallès-Català, T., Pedret, A., Ribes, D., Medina, D., & Traveria, M. (2021). Effects of stress on performance during highly demanding tasks in student pilots. *The International Journal of Aerospace Psychology*, *31*(1), 43–55. https://doi.org/10.1080/24721840.2020.1841564

Vine, S. J., Uiga, L., Lavric, A., Moore, L. J., Tsaneva-Atanasova, K., & Wilson, M. R. (2015). Individual reactions to stress predict performance during a critical aviation incident. *Anxiety, Stress & Coping: An International Journal*, *28*(4), 467–477.

Wang, L., Doucet, L., Waller, M., Sanders, K., & Phillips, S. (2016). A laughing matter: Patterns of laughter and the effectiveness of working dyads. *Organization Science*, *27*(5), 1142–1160. https://doi.org/10.1287/orsc.2016.1082

Xu, S., Martinez, L. R., Van Hoof, H., Eljuri, M. I., & Arciniegas, L. (2016). Fluctuating emotions: Relating emotional variability and job satisfaction. *Journal of Applied Social Psychology*, *46*(11), 617–626. https://doi.org/10.1111/jasp.12390

Zheng, J., Lajoie, S., & Li, S. (2023b). Emotions in self-regulated learning: A critical literature review and meta-analysis. *Frontiers in Psychology*, *14,* 1137010. https://www.frontiersin.org/articles/10.3389/fpsyg.2023.1137010

Zheng, J., Li, S., & Lajoie, S. P. (2023a). A review of measurements and techniques to study emotion dynamics in learning. In V. Kovanovic, R. Azevedo, D. C. Gibson, & D. lfenthaler (Eds.), *Unobtrusive observations of learning in digital environments* (pp. 7–29). Springer. https://doi.org/10.1007/978-3-031-30992-2_2

Bibliography

Ahn, B., & Harley, J. M. (2020). Facial expressions when learning with a queer history app: Application of the control value theory of achievement emotions. *British Journal of Educational Technology*, *51*(5), 1563–1576. https://doi.org/10.1111/bjet.12989

Ahn, B., Maurice-Ventouris, M., Bilgic, E., Yang, A., Lau, C. H.-H., Peters, H., Li, K., Chang-Ou, D., & Harley, J. M. (2023). A scoping review of emotions and related constructs in simulation-based education research articles. *Advances in Simulation*, *8*(1), 22. https://doi.org/10.1186/s41077-023-00258-z

Artino, A. R., Holmboe, E. S., & Durning, S. J. (2012). Control-value theory: Using achievement emotions to improve understanding of motivation, learning, and performance in medical education: AMEE Guide No. 64. *Medical Teacher*, *34*(3), e148–e160. https://doi.org/10.3109/0142159X.2012.651515

Allsop, J., & Gray, R. (2014). Flying under pressure: Effects of anxiety on attention and gaze behavior in aviation. *Journal of Applied Research in Memory and Cognition*, *3*(2), 63–71. https://doi.org/10.1016/j.jarmac.2014.04.010

Althubaiti, A. (2016). Information bias in health research: Definition, pitfalls, and adjustment methods. *Journal of Multidisciplinary Healthcare*, *9*, 211–217. https://doi.org/10.2147/JMDH.S104807

American Psychological Association. (n.d.). Affect. In *APA dictionary of psychology*. Retrieved December 5, 2023, from https://dictionary.apa.org/affect

Artino, A. R., Holmboe, E. S., & Durning, S. J. (2012). Control-value theory: Using achievement emotions to improve understanding of motivation, learning, and performance in medical education: AMEE Guide No. 64. *Medical Teacher*, *34*(3), e148–e160. https://doi.org/10.3109/0142159X.2012.651515

Artino, A. R., & Jones, K. D. (2012). Exploring the complex relations between achievement emotions and self-regulated learning behaviors in online learning. *The Internet and Higher Education*, *15*(3), 170–175. https://doi.org/10.1016/j.iheduc.2012.01.006

Artino, A. R., Rochelle, J. S. L., & Durning, S. J. (2010). Second-year medical students' motivational beliefs, emotions, and achievement. *Medical Education*, *44*(12), 1203–1212. https://doi.org/10.1111/j.1365-2923.2010.03712.x

Azher, S., Cervantes, A., Marchionni, C., Grewal, K., Marchand, H., & Harley, J. M. (2023). Virtual simulation in nursing education: Headset virtual reality and screen-based virtual simulation offer a comparable experience. *Clinical Simulation in Nursing*, *79*, 61–74. https://doi.org/10.1016/j.ecns.2023.02.009

Bailen, N. H., Green, L. M., & Thompson, R. J. (2019). Understanding emotion in adolescents: A review of emotional frequency, intensity, instability, and clarity. *Emotion Review*, *11*(1), 63–73. https://doi.org/10.1177/1754073918768878

Barrett, L. F., Adolphs, R., Marsella, S., Martinez, A. M., & Pollak, S. D. (2019). Emotional expressions reconsidered: Challenges to inferring emotion from human facial movements. *Psychological Science in the Public Interest*, *20*(1), 1–68. https://doi.org/10.1177/1529100619832930

Baumann, M. R., Gohm, C. L., & Bonner, B. L. (2011). Phased training for high-reliability occupations: Live-fire exercises for civilian firefighters. *Human Factors*, *53*(5), 548–557. https://doi.org/10.1177/0018720811418224

Bent, J., & Chan, K. (2010). Flight training and simulation as safety generators. In E. Salas & D. Maurino (Eds.), *Human factors in aviation* (2nd ed., pp. 293–334). Academic Press. https://doi.org/10.1016/B978-0-12-374518-7.00010-9

Berner, E. S., & Graber, M. L. (2008). Overconfidence as a cause of diagnostic error in medicine. *The American Journal of Medicine*, *121*(5 Suppl), S2-23. https://doi.org/10.1016/j.amjmed.2008.01.001

Bogaisky, J. (2023, April 10). *The pilot shortage is playing havoc with air travel. Here are some remedies.* Forbes. https://www.forbes.com/sites/jeremybogaisky/2023/04/10/the-pilot-shortage-is-playing-in-havoc-with-air-travel-here-are-some-remedies/

Boucsein, W., Fowles, D. C., Grimnes, S., Ben-Shakhar, G., Walton, T. Roth, Dawson, M. E., & Filion, D. L. (2012). Publication recommendations for electrodermal measurements. *Psychophysiology*, *49*(8), 1017–1034. https://doi.org/10.1111/j.1469-8986.2012.01384.x

Braithwaite, J. J., Watson, D. G., Jones, R., & Rowe, M. (2015). *A guide for analysing electrodermal activity (EDA) and skin conductance responses (SCRs) for psychological experiments*. SAAL, Behavioural Brain Sciences Centre, University of Birminghan. https://www.birmingham.ac.uk/Documents/college-les/psych/saal/guide-electrodermal-activity.pdf

Canadian Aviation Regulations. (2022, December 21). *Commercial air services*. Justice laws website - Government of Canada. https://laws-lois.justice.gc.ca/eng/regulations/sor-96-433/page-83.html

Cao, X., MacNaughton, P., Cadet, L. R., Cedeno-Laurent, J. G., Flanigan, S., Vallarino, J., Donnelly-McLay, D., Christiani, D. C., Spengler, J. D., & Allen, J. G. (2019). Heart rate variability and performance of commercial airline pilots during flight simulations.

International Journal of Environmental Research and Public Health, *16*(2), 237. https://doi.org/10.3390/ijerph16020237

Carden, S. W., Holtzman, N. S., & Strube, M. J. (2017). CAHOST: An excel workbook for facilitating the Johnson-Neyman technique for two-way interactions in multiple regression. *Frontiers in Psychology*, *8*, 1293. https://doi.org/10.3389/fpsyg.2017.01293

Causse, M., Mouratille, D., Rouillard, Y., El Yagoubi, R., Matton, N., & Hidalgo-Muñoz, A. (2024). How a pilot's brain copes with stress and mental load? Insights from the executive control network. *Behavioural Brain Research*, *456*, 114698. https://doi.org/10.1016/j.bbr.2023.114698

Code of Federal Regulations. (2023, April 19). *14 CFR 91.1059 -- Flight time limitations and rest requirements: One or two pilot crews.* Electronic code of federal regulations. https://www.ecfr.gov/current/title-14/chapter-I/subchapter-F/part-91/subpart-K/subject-group-ECFRc17623c0e0be17e/section-91.1059

Cohen, J. (1988). The Analysis of Variance. In *Statistical Power Analysis for the Behavioral Sciences* (2nd ed.). Routledge.

Cootes, T. F., & Taylor, C. J. (2001). Statistical models of appearance for medical image analysis and computer vision. *Medical Imaging 2001: Image Processing*, *4322*, 236–248. https://doi.org/10.1117/12.431093

Cross, J. I., Boag-Hodgson, C. C., & Mavin, T. J. (2023). Measuring presence and situational awareness in a virtual reality flight simulator. *Aviation Psychology and Applied Human Factors*. https://doi.org/10.1027/2192-0923/a000250

Dai, J., Wang, H., Yang, L., & Wen, Z. (2019). Emotional intelligence and emotional state effects on simulated flight performance. *Aerospace Medicine and Human Performance*, *90*(2), 101–108. https://doi.org/10.3357/AMHP.5184.2019

Damasio, A. R. (2005). *Descartes' error: Emotion, reason, and the human brain*. Penguin Books.

DeMaria, S., Bryson, E. O., Mooney, T. J., Silverstein, J. H., Reich, D. L., Bodian, C., & Levine, A. I. (2010). Adding emotional stressors to training in simulated cardiopulmonary arrest enhances participant performance. *Medical Education*, *44*(10), 1006–1015. https://doi.org/10.1111/j.1365-2923.2010.03775.x

Dismukes, R. K. (2010). Understanding and analyzing human error in real-world operations. In E. Salas & D. Maurino (Eds.), *Human factors in aviation* (2nd ed., pp. 335–374). Academic Press. https://doi.org/10.1016/B978-0-12-374518-7.00011-0

D'Mello, S., & Graesser, A. (2012). Dynamics of affective states during complex learning. *Learning and Instruction*, *22*(2), 145–157. https://doi.org/10.1016/j.learninstruc.2011.10.001

D'Mello, S., Lehman, B., & Person, N. (2010). Monitoring affect states during effortful problem solving activities. *International Journal of Artificial Intelligence in Education*, *20*(4), 361–389. https://doi.org/10.3233/JAI-2010-012

Drinkwater, B. L., Flint, M. M., Clel, & Troy S. (1968). Somatic responses and performance levels during anticipatory physical-threat stress. *Perceptual and Motor Skills*, *27*(2), 539–552. https://doi.org/10.2466/pms.1968.27.2.539

Duffy, M. C., Lajoie, S., & Lachapelle, K. (2016). Measuring emotions in medical education: Methodological and technological advances within authentic medical learning environments. In S. Bridges, L. K. Chan, & C. E. Hmelo-Silver (Eds.), *Educational* *technologies in medical and health sciences education* (pp. 181–213). Springer. https://doi.org/10.1007/978-3-319-08275-2_10

Duffy, M. C., Lajoie, S. P., Pekrun, R., & Lachapelle, K. (2018). Emotions in medical education: Examining the validity of the Medical Emotion Scale (MES) across authentic medical learning environments. *Learning and Instruction*, 101150. https://doi.org/10.1016/j.learninstruc.2018.07.001

Ekman, P. (1992). An argument for basic emotions. *Cognition and Emotion*, *6*(3–4), 169–200. https://doi.org/10.1080/02699939208411068

Ekman, P. (2020). *What are emotions?* Paul Ekman Group. https://www.paulekman.com/universal-emotions/

Ekman, P., & Rosenberg, E. L. (Eds.). (2005). *What the face reveals: Basic and applied studies of spontaneous expression using the facial action coding system (FACS)* (2nd ed.). Oxford University Press. 10.1093/acprof:oso/9780195179644.001.0001

Endsley, M. R. (2000). Theoretical underpinnings of situation awareness: A critical review. In M. R. Endsley & D. J. Garland (Eds.), *Situation Awareness Analysis and Measurement* (pp. 3-28). Lawrence Erlbaum Associates.

Ericsson, K. A. (2006). The influence of experience and deliberate practice on the development of superior expert performance. In K. A. Ericsson, N. Charness, P. J. Feltovich, & R. R. Hoffman (Eds.), *The Cambridge handbook of expertise and expert performance* (pp. 683–704). Cambridge University Press. https://doi.org/10.1017/CBO9780511816796.038

European Commission. (2013, October 4). *Pilot and crew fatigue—Frequently asked questions*. Directorate-General for Communication, European Commission. https://ec.europa.eu/commission/presscorner/detail/en/MEMO_13_854

Eysenck, M. W., Derakshan, N., Santos, R., & Calvo, M. G. (2007). Anxiety and cognitive performance: Attentional control theory. *Emotion*, *7*(2), 336–353. https://doi.org/10.1037/1528-3542.7.2.336

FAA Department of Transportation. (2003, September 17). *14 CFR 91.1059 -- Flight time limitations and rest requirements: One or two pilot crews.* Electronic Code of Federal Regulations. https://www.ecfr.gov/current/title-14/chapter-I/subchapter-F/part-91/subpart-K/subject-group-ECFRc17623c0e0be17e/section-91.1059

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*(2), 175–191. https://doi.org/10.3758/bf03193146

Fink, G. (2016). Stress, definitions, mechanisms, and effects outlined: Lessons from anxiety. In G. Fink (Ed.), *Stress: Concepts, cognition, emotion, and behavior* (pp. 3–11). Academic Press. https://doi.org/10.1016/B978-0-12-800951-2.00001-7

Gaba, D. M. (2004). The future vision of simulation in health care. *Quality & Safety in Health Care*, *13 Suppl 1*(Suppl 1), i2-10. https://doi.org/10.1136/qhc.13.suppl_1.i2

Gaetan, S., Dousset, E., Marqueste, T., Bringoux, L., Bourdin, C., Vercher, J.-L., & Besson, P. (2015). Cognitive workload and psychophysiological parameters during multitask activity in helicopter pilots. *Aerospace Medicine and Human Performance*, *86*(12), 1052–1057. https://doi.org/10.3357/AMHP.4228.2015

Gao, Z., & Xiang, P. (2008). College students' motivation toward weight training: An application of expectancy-value model. *Journal of Teaching in Physical Education*, *27*(3), 399–415. https://doi.org/10.1123/jtpe.27.3.399

Goetz, T., & Hall, N. C. (2014). Academic boredom. In R. Pekrun & L. Linnenbrink-Garcia (Eds.), *International handbook of emotions in education* (pp. 311–330). Routledge.

Gordon, J. A., Hayden, E. M., Ahmed, R. A., Pawlowski, J. B., Khoury, K. N., & Oriol, N. E. (2010). Early bedside care during preclinical medical education: Can technology-enhanced patient simulation advance the Flexnerian ideal? *Academic Medicine*, *85*(2), 370–377. https://doi.org/10.1097/ACM.0b013e3181c88d74

Gray, R., Gaska, J., & Winterbottom, M. (2016). Relationship between sustained, orientated, divided, and selective attention and simulated aviation performance: Training & pressure effects. *Journal of Applied Research in Memory and Cognition*, *5*(1), 34–42. https://doi.org/10.1016/j.jarmac.2015.11.005

Gross, J. J. (2015). Emotion regulation: Current status and future prospects. *Psychological Inquiry*, *26*(1), 1–26. https://doi.org/10.1080/1047840X.2014.940781

Gruber, J., Kogan, A., Quoidbach, J., & Mauss, I. B. (2013). Happiness is best kept stable: Positive emotion variability is associated with poorer psychological health. *Emotion*, *13*(1), 1–6. https://doi.org/10.1037/a0030262

Hamman, W. R. (2004). Will simulation fly in medicine as it has in aviation? *Quality & Safety in Health Care*, *13*(5), 397–399. https://doi.org/10.1136/qhc.13.5.397

Han, S.-Y., Kwak, N.-S., Oh, T., & Lee, S.-W. (2020). Classification of pilots' mental states using a multimodal deep learning network. *Biocybernetics and Biomedical Engineering*, *40*(1), 324–336. https://doi.org/10.1016/j.bbe.2019.12.002

Harley, J. M. (2016). Measuring emotions: A survey of cutting edge methodologies used in computer-based learning environment research. In S. Y. Tettegah & M. Gartmeier (Eds.), *Emotions, technology, design, and learning* (pp. 89–114). Academic Press. https://doi.org/10.1016/B978-0-12-801856-9.00005-0

Harley, J. M., & Azevedo, R. (2014). Toward a feature-driven understanding of students' emotions during interactions with agent-based learning environments: A selective review. *International Journal of Gaming and Computer-Mediated Simulations*, *6*(3), 17–34. https://doi.org/10.4018/ijgcms.2014070102

Harley, J. M., Bilgic, E., Lau, C. H. H., Gorgy, A., Marchand, H., Lajoie, S. P., Lavoie-Tremblay, M., & Fried, G. M. (2023). Nursing students reported more positive emotions about training during Coronavirus disease 2019 (COVID-19) after using a virtual simulation paired with an in-person simulation. *Clinical Simulation in Nursing*, *81*, 101420. https://doi.org/10.1016/j.ecns.2023.04.006

Harley, J. M., Bouchet, F., & Azevedo, R. (2012). Measuring learners' co-occurring emotional responses during their interaction with a pedagogical agent in MetaTutor. In S. A. Cerri, W. J. Clancey, G. Papadourakis, & K. Panourgia (Eds.), *Intelligent tutoring systems* (pp. 40–45). Springer. https://doi.org/10.1007/978-3-642-30950-2_5

Harley, J. M., Bouchet, F., & Azevedo, R. (2013). Aligning and comparing data on emotions experienced during learning with MetaTutor. In H. C. Lane, K. Yacef, J. Mostow, & P. Pavlik (Eds.), *Artificial intelligence in education* (pp. 61–70). Springer. https://doi.org/10.1007/978-3-642-39112-5_7

Harley, J. M., Bouchet, F., Hussain, M. S., Azevedo, R., & Calvo, R. (2015). A multi-componential analysis of emotions during complex learning with an intelligent multi-agent system. *Computers in Human Behavior*, *48*, 615–625. https://doi.org/10.1016/j.chb.2015.02.013

Harley, J. M., Jarrell, A., & Lajoie, S. P. (2019b). Emotion regulation tendencies, achievement emotions, and physiological arousal in a medical diagnostic reasoning simulation. *Instructional Science*, *47*(2), 151–180. https://doi.org/10.1007/s11251-018-09480-z

Harley, J. M., Lajoie, S. P., Frasson, C., & Hall, N. C. (2017). Developing Emotion-Aware, Advanced Learning Technologies: A Taxonomy of Approaches and Features. *International Journal of Artificial Intelligence in Education*, *27*(2), 268–297. https://doi.org/10.1007/s40593-016-0126-8

Harley, J. M., Pekrun, R., Taxer, J. L., & Gross, J. J. (2019a). Emotion regulation in achievement situations: An integrated model. *Educational Psychologist*, *54*(2), 106–126. https://doi.org/10.1080/00461520.2019.1587297

Hart, S. G. (2006). Nasa-Task Load Index (NASA-TLX): 20 years later. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *50*(9), 904–908. https://doi.org/10.1177/154193120605000909

Hart, S. G., & Bortolussi, M. R. (1984). Pilot errors as a source of workload. *Special Issue: Aviation Psychology*, *26*(5), 545–556.

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In P. A. Hancock & N. Meshkati (Eds.), *Human mental workload* (Vol. 52, pp. 139–183). Elsevier Science. https://doi.org/10.1016/S0166-4115(08)62386-9

Hayes, A. F. (2022). Moderation analysis. In *Introduction to Mediation, Moderation, and Conditional Process Analysis: A Regression-Based Approach* (pp. 231–281). Guilford Press.

Helmreich, R. L. (2010). Foreword. In E. Salas & D. Maurino (Eds.), *Human factors in aviation* (2nd ed., pp. xi–xii). Academic Press. https://doi.org/10.1016/B978-0-12-374518-7.00024-9

Herrera-Aliaga, E., & Estrada, L. D. (2022). Trends and innovations of simulation for twenty first century medical education. *Frontiers in Public Health*, *10*, 619769. https://doi.org/10.3389/fpubh.2022.619769

Hidalgo-Munoz, A. R., Mouratille, D., Matton, N., Causse, M., Rouillard, Y., & El-Yagoubi, R. (2018). Cardiovascular correlates of emotional state, cognitive workload and time-on-task effect during a realistic flight simulation. *International Journal of Psychophysiology*, *128*, 62–69. https://doi.org/10.1016/j.ijpsycho.2018.04.002

Hodhod, R., Fleenor, H., & Nabi, S. (2014). Adaptive augmented reality serious game to foster problem solving skills. *Workshop Proceedings of the 10th International Conference on Intelligent Environments*, 273–284. https://doi.org/10.3233/978-1-61499-411-4-273

Hou, C., Zhu, G., & Yang, Y. (2023, June 10). *Beyond the cognitive dimension: Emotion patterns in productive and improvable knowledge building discourse*. Annual Meeting of the International Society of the Learning Sciences. https://doi.org/10.22318/cscl2023.333889

Houben, M., Van Den Noortgate, W., & Kuppens, P. (2015). The relation between short-term emotion dynamics and psychological well-being: A meta-analysis. *Psychological Bulletin*, *141*(4), 901–930. https://doi.org/10.1037/a0038822

Hunter, J. D. (2007). Matplotlib: A 2D graphics environment. *Computing in Science & Engineering*, *9*(3), 90–95. https://doi.org/10.1109/MCSE.2007.55

Jack, R. E., Garrod, O. G. B., & Schyns, P. G. (2014). Dynamic facial expressions of emotion transmit an evolving hierarchy of signals over time. *Current Biology: CB*, *24*(2), 187–192. https://doi.org/10.1016/j.cub.2013.11.064

Jarrell, A. (2015). *The emotional twists and turns of problem solving: An examination of learners' behavioral, phsyiological and experiential emotion responses to unexpected events during problem solving* [Masters thesis, McGill University]. https://escholarship.mcgill.ca/concern/theses/3197xp56q

Jarrell, A., Harley, J. M., Lajoie, S., & Naismith, L. (2017). Success, failure and emotions: Examining the relationship between performance feedback and emotions in diagnostic reasoning. *Educational Technology Research and Development*, *65*(5), 1263–1284. https://doi.org/10.1007/s11423-017-9521-6

Jarrell, A., & Lajoie, S. P. (2017). The regulation of achievements emotions: Implications for research and practice. *Canadian Psychology/Psychologie Canadienne*, *58*(3), 276–287. https://doi.org/10.1037/cap0000119

Jennings, S., Law, A., Bourgon, A., & Grenier, H. (2024, January). *Novel methodology for comparing student pilot flight performance with instructor pilot expert ratings*. AIAA SCITECH 2024 Forum, Orlando, Florida.

Jorna, P. G. (1993). Heart rate and workload variations in actual and simulated flight. *Special Issue: Psychophysiological Measures in Transport Operations*, *36*(9), 1043–1054. https://doi.org/10.1080/00140139308967976 Full text linksCite

Kallus, K. W., & Tropper, K. (2004). Evaluation of a spatial disorientation simulator training for jet pilots. *International Journal of Applied Aviation Studies*, *4*(1), 45–55. https://doi.org/10.1080/10508414.2011.556458

Kang, Y., Yun, M. H., & Kim, S. (2020). Verbal reports' influence on pilot flight performance and mental stress under spatial disorientation. *Aerospace Medicine and Human Performance*, *91*(12), 948–955. https://doi.org/10.3357/AMHP.5620.2020

Karaca, Y., & Moonis, M. (2022). Shannon entropy-based complexity quantification of nonlinear stochastic process: Diagnostic and predictive spatiotemporal uncertainty of multiple sclerosis subgroups. In Y. Karaca, D. Baleanu, Y.-D. Zhang, O. Gervasi, & M. Moonis (Eds.), *Multi-Chaos, Fractal and Multi-Fractional Artificial Intelligence of Different Complex Systems* (pp. 231–245). Academic Press. https://doi.org/10.1016/B978-0-323-90032-4.00018-3

Kashdan, T. B., & Rottenberg, J. (2010). Psychological flexibility as a fundamental aspect of health. *Clinical Psychology Review*, *30*(7), 865–878. https://doi.org/10.1016/j.cpr.2010.03.001

Kinney, L., & O'Hare, D. (2020). Responding to an unexpected in-flight event: Physiological arousal, information processing, and performance. *Human Factors*, *62*(5), 737–750. https://doi.org/10.1177/0018720819854830

Koglbauer, I., Kallus, K., Braunstingl, R., & Boucsein, W. (2011). Recovery training in simulator improves performance and psychophysiological state of pilots during simulated and real visual flight rules flight. *The International Journal of Aviation Psychology*, *21*(4), 307–324. https://doi.org/10.1080/10508414.2011.606741

Krahenbuhl, G. S., Darst, P. W., Marett, J. R., Reuther, L. C., Constable, S. H., Swinford, M. E., & Reid, G. B. (1981). Instructor pilot teaching behavior and student pilot stress in flight training. *Aviation, Space, and Environmental Medicine*, *52*(10), 594–597.

Krahenbuhl, G. S., Marett, J. R., & Reid, G. B. (1978). Task-specific simulator pretraining and in-flight stress of student pilots. *Aviation, Space, and Environmental Medicine*, *49*, 1107–1110.

Krone, T., Albers, C. J., Kuppens, P., & Timmerman, M. E. (2018). A multivariate statistical model for emotion dynamics. *Emotion*, *18*(5), 739–754. https://doi.org/10.1037/emo0000384

Kuppens, P. (2015). It's about time: A special section on affect dynamics. *Emotion Review*, *7*(4), 297–300. https://doi.org/10.1177/1754073915590947

Lajoie, S. P. (2017). Learning science applications for research in medicine. In L. Lin & J. M. Spector (Eds.), *The sciences of learning and instructional design* (pp. 108–118). Routledge. https://doi.org/10.4324/9781315684444-8

Lajoie, S. P. (2021). Multimedia learning with simulations. In L. Fiorella & R. E. Mayer (Eds.), *The Cambridge handbook of multimedia learning* (3rd ed., pp. 461–471). Cambridge University Press. https://doi.org/10.1017/9781108894333.048

Lajoie, S. P., & Azevedo, R. (2006). Teaching and learning in technology-rich environments. In P. A. Alexander & P. H. Winne (Eds.), *Handbook of educational psychology* (pp. 803–821). Lawrence Erlbaum Associates Publishers.

Lajoie, S. P., & Gube, M. (2018). Adaptive expertise in medical education: Accelerating learning trajectories by fostering self-regulated learning. *Medical Teacher*, *40*(8), 809–812. https://doi.org/10.1080/0142159X.2018.1485886

Lajoie, S. P., Li, S., & Zheng, J. (2023). The functional roles of metacognitive judgement and emotion in predicting clinical reasoning performance with a computer simulated environment. *Interactive Learning Environments*, *31*(6), 3464–3475. https://doi.org/10.1080/10494820.2021.1931347

Lajoie, S. P., Naismith, L., Poitras, E., Hong, Y.-J., Cruz-Panesso, I., Ranellucci, J., Mamane, S., & Wiseman, J. (2013). Technology-rich tools to support self-regulated learning and performance in medicine. In R. Azevedo & V. Aleven (Eds.), *International handbook of metacognition and learning technologies* (pp. 229–242). Springer. https://doi.org/10.1007/978-1-4419-5546-3_16

Lajoie, S. P., Pekrun, R., Azevedo, R., & Leighton, J. P. (2020). Understanding and measuring emotions in technology-rich learning environments. *Learning and Instruction*, *70*, 101272. https://doi.org/10.1016/j.learninstruc.2019.101272

Lajoie, S. P., & Poitras, E. (2023). Technology-rich learning environments: Theories and methodologies for understanding solo and group learning. In P. A. Schutz & K. R. Muis (Eds.), *Handbook of educational psychology* (4th ed., pp. 630–653). Routledge. https://doi.org/10.4324/9780429433726

Lajoie, S. P., Zheng, J., Li, S., Jarrell, A., & Gube, M. (2021). Examining the interplay of affect and self-regulation in the context of clinical reasoning. *Learning and Instruction*, *72*, 101219. https://doi.org/10.1016/j.learninstruc.2019.101219

Laminar Research. (2022). *X-Plane 11 Desktop manual*. https://www.x-plane.com/manuals/desktop/

Landman, A., Groen, E. L., van Paassen, M. M., Bronkhorst, A. W., Mulder, M., & Mulder, M. (2017). Dealing with unexpected events on the flight deck: A conceptual model of startle and surprise. *Human Factors*, *59*(8), 1161–1172. https://doi.org/10.1177/0018720817723428

Landman, A., van Middelaar, S. H., Groen, E. L., van Paassen, M. M., Bronkhorst, A. W., & Mulder, M. (2020). The effectiveness of a mnemonic-type startle and surprise management procedure for pilots. *The International Journal of Aerospace Psychology*, *30*(3–4), 104–118. https://doi.org/10.1080/24721840.2020.1763798

Lazarus, R. S., PhD, & Folkman, S., PhD. (1984). *Stress, appraisal, and coping*. Springer. 10.1007/978-1-4419-1005-9_215

Lesne, A. (2014). Shannon entropy: A rigorous notion at the crossroads between probability, information theory, dynamical systems and statistical physics. *Mathematical Structures in Computer Science*, *24*(3), e240311. https://doi.org/10.1017/S0960129512000783

Lewinski, P., den Uyl, T. M., & Butler, C. (2014). Automated facial coding: Validation of basic emotions and FACS AUs in FaceReader. *Journal of Neuroscience, Psychology, and Economics*, *7*(4), 227–236. https://doi.org/10.1037/npe0000028

Li, S., Zheng, J., & Lajoie, S. P. (2021a). The frequency of emotions and emotion variability in self-regulated learning: What matters to task performance? *Frontline Learning Research*, *9*(4), 76–91. https://doi.org/10.14786/flr.v9i4.901

Li, S., Zheng, J., Lajoie, S. P., & Wiseman, J. (2021b). Examining the relationship between emotion variability, self-regulated learning, and task performance in an intelligent tutoring system. *Educational Technology Research and Development*, *69*(2), 673–692. https://doi.org/10.1007/s11423-021-09980-9

Li, T. (2020). *Multimodal affect assessment in aviation training* [Masters thesis, McGill University]. https://escholarship.mcgill.ca/concern/theses/6d5702130

Li, T., & Lajoie, S. (2021). Predicting aviation training performance with multimodal affective inferences. *International Journal of Training and Development*, *25*(3), 301–315. https://doi.org/10.1111/ijtd.12232

Lobczowski, N. G. (2022). Capturing the formation and regulation of emotions in collaborative learning: The FRECL coding procedure. *Frontiers in Psychology*, *13*, 846811. https://doi.org/10.3389/fpsyg.2022.846811

Loijens, L., & Krips, O. (2021). *FaceReader methodology note*. Noldus Information Technology.

Loijens, L., Krips, O., van Kuilenburg, H., den Uyl, M., & Ivan, P. (2015). *FaceReader version 6.1 Reference manual*. Noldus Information Technology.

Makowski, D., Pham, T., Lau, Z. J., Brammer, J. C., Lespinasse, F., Pham, H., Schölzel, C., & Chen, S. H. A. (2021). NeuroKit2: A Python toolbox for neurophysiological signal processing. *Behavior Research Methods*, *53*(4), 1689–1696. https://doi.org/10.3758/s13428-020-01516-y

Makransky, G., Andreasen, N. K., Baceviciute, S., & Mayer, R. E. (2020). Immersive virtual reality increases liking but not learning with a science simulation and generative learning strategies promote learning in immersive virtual reality. *Journal of Educational Psychology*, *113*, 719–735. https://doi.org/10.1037/edu0000473

Marques, E., Carim, G., Campbell, C., & Lohmann, G. (2023). Ab initio flight training: A Systematic literature review. *The International Journal of Aerospace Psychology*, *33*(2), 99–119. https://doi.org/10.1080/24721840.2022.2162405

Martín-Gutiérrez, J., Mora, C. E., Añorbe-Díaz, B., & González-Marrero, A. (2017). Virtual Technologies Trends in Education. *Eurasia Journal of Mathematics, Science and Technology Education*, *13*(2), 469–486. https://doi.org/10.12973/eurasia.2017.00626a

Martin, W. L., Murray, P. S., Bates, P. R., & Lee, P. S. Y. (2015). Fear-potentiated startle: A review from an aviation perspective. *The International Journal of Aviation Psychology*, *25*(2), 97–107. https://doi.org/10.1080/10508414.2015.1128293

Martins, A. P. G. (2016). A review of important cognitive concepts in aviation. *Aviation*, *20*(2), 65–84. https://doi.org/10.3846/16487788.2016.1196559

Mathavara, K., & Ramachandran, G. (2022). Role of human factors in preventing aviation accidents: An insight. In Z. A. Ali & D. Cvetković (Eds.), *Aeronautics—New advances*. IntechOpen. https://doi.org/10.5772/intechopen.106899

McClernon, C. K., McCauley, M. E., O'Connor, P. E., & Warm, J. S. (2011). Stress training improves performance during a stressful flight. *Human Factors*, *53*(3), 207–218. https://doi.org/10.1177/0018720811405317

McClernon, C. K., & Miller, J. C. (2011). Variance as a measure of performance in an aviation context. *The International Journal of Aviation Psychology*, *21*(4), 397–412. https://doi.org/10.1080/10508414.2011.606765

Merriam-Webster. (2023). *Definition of simulator*. https://www.merriam-webster.com/dictionary/simulator

Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. G. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *BMJ*, *339*, b2535. https://doi.org/10.1136/bmj.b2535

Morris, C. S., Hancock, P. A., & Shirkey, E. C. (2004). Motivational effects of adding context relevant stress in PC-based game training. *Military Psychology*, *16*(2), 135–147. https://doi.org/10.1207/S15327876MP1602_4

Muñoz, J. E., Montoya, M. F., & Boger, J. (2022). From exergames to immersive virtual reality systems: Serious games for supporting older adults. In M.-A. Choukou & S. Syed-Abdul (Eds.), *Smart home technologies and services for geriatric rehabilitation* (pp. 141–204). Academic Press. https://doi.org/10.1016/B978-0-323-85173-2.00011-4

Murray, G., Heilakka, R., & Medland, A. (2022, November 1). *The pilot shortage will only get worse as the decade progresses*. BRINK – Conversations and Insights on Global Business. https://www.brinknews.com/the-pilot-shortage-will-only-get-worse-as-the-decade-progresses/

Naismith, L. M., & Lajoie, S. P. (2018). Motivation and emotion predict medical students' attention to computer-based feedback. *Advances in Health Sciences Education*, *23*(3), 465–485. https://doi.org/10.1007/s10459-017-9806-x

Noldus. (2015). *FaceReader version 6.1 Reference manual*. Noldus Information Technology.

Nomura, O., Wiseman, J., Sunohara, M., Akatsu, H., & Lajoie, S. P. (2021). Japanese medical learners' achievement emotions: Accounting for culture in translating Western medical educational theories and instruments into an asian context. *Advances in Health Sciences Education*, *26*(4), 1255–1276. https://doi.org/10.1007/s10459-021-10048-9

Ouzzani, M., Hammady, H., Fedorowicz, Z., & Elmagarmid, A. (2016). Rayyan—A web and mobile app for systematic reviews. *Systematic Reviews*, *5*(1), 210. https://doi.org/10.1186/s13643-016-0384-4

Page, M. J., Moher, D., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., … McKenzie, J. E. (2021). PRISMA 2020 explanation and elaboration: Updated guidance and exemplars for reporting systematic reviews. *BMJ*, *372*, n160. https://doi.org/10.1136/bmj.n160

Parong, J., & Mayer, R. E. (2020). Cognitive and affective processes for learning science in immersive virtual reality. *Journal of Computer Assisted Learning*, *1*(16). https://doi.org/10.1111/jcal.12482

Pekrun, R. (2019). Achievement emotions: A control-value theory perspective. In R. Patulny, A. Bellocchi, R. Olson, S. Khorana, J. McKenzie, & M. Peterie (Eds.), *Emotions in late modernity* (pp. 142-157). Routledge.

Pekrun, R., & Linnenbrink-Garcia, L. (2014). Introduction to emotions in education. In Pekrun, Reinhard & Linnenbrink-Garcia, Lisa (Eds.), *International handbook of emotions in education* (pp. 1–10). Routledge.

Pekrun, R., & Perry, R. P. (2014). Control-value theory of achievement emotions. In R. Pekrun & L. Linnenbrink-Garcia (Eds.), *International handbook of emotions in education* (pp. 120–141). Routledge.

Perry, R. P., Hladkyj, S., Pekrun, R. H., & Pelletier, S. T. (2001). Academic control and action control in the achievement of college students: A longitudinal field study. *Journal of Educational Psychology*, *93*, 776–789. https://doi.org/10.1037/0022-0663.93.4.776

Pilote, B., & Chiniara, G. (2019). The many faces of simulation. In G. Chiniara (Ed.), *Clinical simulation* (2nd ed., pp. 17–32). Academic Press. https://doi.org/10.1016/B978-0-12-815657-5.00002-4

Roh, Y. S., & Jang, K. I. (2017). Survey of factors influencing learner engagement with simulation debriefing among nursing students. *Nursing & Health Sciences*, *19*(4), 485–491. https://doi.org/10.1111/nhs.12371

Rosa, E., Gronkvist, M., Kolegard, R., Dahlstrom, N., Knez, I., Ljung, R., & Willander, J. (2021). Fatigue, emotion, and cognitive performance in simulated long-duration, single-piloted flight missions. *Aerospace Medicine and Human Performance*, *92*(9), 710–719. https://doi.org/10.3357/AMHP.5798.2021

Rosa, E., Lyskov, E., Gronkvist, M., Kolegard, R., Dahlstrom, N., Knez, I., Ljung, R., & Willander, J. (2022). Cognitive performance, fatigue, emotional, and physiological strains in simulated long-duration flight missions. *Military Psychology*, *34*(2), 224–236. https://doi.org/10.1080/08995605.2021.1989236

Rudland, J. R., Golding, C., & Wilkinson, T. J. (2020). The stress paradox: How stress can be good for learning. *Medical Education*, *54*(1), 40–45. https://doi.org/10.1111/medu.13830

Ruiz Segura, A. (2020). *Examining the interaction between clinical reasoning and emotion on medical students' diagnostic efficiency* [Masters thesis, McGill University]. https://escholarship.mcgill.ca/concern/theses/5h73q1579

Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology*, *39*(6), 1161–1178. https://doi.org/10.1037/h0077714

Russell, J. A. (2003). Core affect and the psychological construction of emotion. *Psychological Review*, *110*(1), 145–172. https://doi.org/10.1037/0033-295X.110.1.145

Salas, E., Maurino, D., & Curtis, M. (2010). Human factors in aviation: An overview. In E. Salas & D. Maurino (Eds.), *Human factors in aviation* (2nd ed., pp. 3–19). Academic Press. https://doi.org/10.1016/B978-0-12-374518-7.00001-8

Sassenus, S., Van den Bossche, P., & Poels, K. (2022). When stress becomes shared: Exploring the emergence of team stress. *Cognition, Technology & Work*, *24*, 537–556. https://doi.org/10.1007/s10111-022-00698-z

Sawyer, R. K., & Greeno, J. G. (2012). Situativity and learning. In P. Robbins & M. Aydede (Eds.), *The Cambridge handbook of situated cognition* (pp. 347–367). https://doi.org/10.1017/CBO9780511816826.018

Shannon, C. E. (1948). A mathematical theory of communication. *The Bell System Technical Journal*, *27*(3), 379–423. https://doi.org/10.1002/j.1538-7305.1948.tb01338.x

Shute, V. J. (2008). Focus on formative feedback. *Review of Educational Research*, *78*(1), 153–189. https://doi.org/10.3102/0034654307313795

Sieberichs, S., & Kluge, A. (2018). Effects of in-flight countermeasures to mitigate fatigue risks in aviation. *Aviation Psychology and Applied Human Factors*, *8*(2), 86–92. https://doi.org/10.1027/2192-0923/a000143

Silva, D. C., Vinhas, V., Reis, L. P., & Oliveira, E. (2009). Biometric emotion assessment and feedback in an immersive digital environment. *International Journal of Social Robotics*, *1*(4), 307–317. https://doi.org/10.1007/s12369-009-0029-z

Skibniewski, F. W., Dziuda, L., Baran, P. M., Krej, M. K., Guzowski, S., Piotrowski, M. A., & Truszczynski, O. E. (2015). Preliminary results of the LF/HF ratio as an indicator for estimating difficulty level of flight tasks. *Aerospace Medicine and Human Performance*, *86*(6), 518–523. https://doi.org/10.3357/AMHP.4087.2015

Smith, R. C., & Melton, C. E. (1978). Effects of ground trainer use on the anxiety of students in private pilot training. *Aviation, Space, and Environmental Medicine*, *49*, 406–408.

Su, C., Li, X., Yang, L., & Zeng, Y. (2022). *Mechanism, measurement, and quantification of stress in decision process: A model based systematic-review protocol* (arXiv:2203.10397). arXiv. https://doi.org/10.48550/arXiv.2203.10397

Thompson, R. J., Mata, J., Jaeggi, S. M., Buschkuehl, M., Jonides, J., & Gotlib, I. H. (2012). The everyday emotional experience of adults with major depressive disorder: Examining emotional instability, inertia, and reactivity. *Journal of Abnormal Psychology*, *121*(4), 819–829. https://doi.org/10.1037/a0027978

Tichon, J. G., Mavin, T., Wallis, G., Visser, T. A. W., & 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.

Tichon, J. G., Wallis, G., Riek, S., & Mavin, T. (2014). Physiological measurement of anxiety to evaluate performance in simulation training. *Cognition, Technology & Work*, *16*(2), 203–210.

Transport Canada. (2019, November 4). *Standard 421—Flight Crew Permits, Licences and Ratings—Canadian Aviation Regulations (CARs)*. AARBH 15472583; AARBH. https://tc.canada.ca/en/corporate-services/acts-regulations/list-regulations/canadian-aviation-regulations-sor-96-433/standards/standard-421-flight-crew-permits-licences-ratings-canadian-aviation-regulations-cars

Tropper, K., Kallus, K. W., & Boucsein, W. (2009). Psychophysiological evaluation of an antidisorientation training for visual flight rules pilots in a moving base simulator. *The International Journal of Aviation Psychology*, *19*(3), 270–286. https://doi.org/10.1080/10508410902983912

Trull, T. J., Solhan, M. B., Tragesser, S. L., Jahng, S., Wood, P. K., Piasecki, T. M., & Watson, D. (2008). Affective instability: Measuring a core feature of borderline personality disorder with ecological momentary assessment. *Journal of Abnormal Psychology*, *117*(3), 647–661. https://doi.org/10.1037/a0012532

Vallès-Català, T., Pedret, A., Ribes, D., Medina, D., & Traveria, M. (2021). Effects of stress on performance during highly demanding tasks in student pilots. *The International Journal of Aerospace Psychology*, *31*(1), 43–55. https://doi.org/10.1080/24721840.2020.1841564

Vine, S. J., Uiga, L., Lavric, A., Moore, L. J., Tsaneva-Atanasova, K., & Wilson, M. R. (2015). Individual reactions to stress predict performance during a critical aviation incident. *Anxiety, Stress & Coping: An International Journal*, *28*(4), 467–477.

Viola, P., & Jones, M. J. (2004). Robust real-time face detection. *International Journal of Computer Vision*, *57*(2), 137–154. https://doi.org/10.1023/B:VISI.0000013087.49260.fb

Wang, L., Doucet, L., Waller, M., Sanders, K., & Phillips, S. (2016). A laughing matter: Patterns of laughter and the effectiveness of working dyads. *Organization Science*, *27*(5), 1142–1160. https://doi.org/10.1287/orsc.2016.1082

Weingarten, N. C. (2005). History of in-flight simulation at general dynamics. *Journal of Aircraft*, *42*(2), 290–298. https://doi.org/10.2514/1.4663

Wigfield, A., & Eccles, J. S. (2000). Expectancy–value theory of achievement motivation. *Contemporary Educational Psychology*, *25*(1), 68–81. https://doi.org/10.1006/ceps.1999.1015

Wingelaar-Jagt, Y. Q., Wingelaar, T. T., Riedel, W. J., & Ramaekers, J. G. (2021). Fatigue in aviation: Safety risks, preventive strategies and pharmacological interventions. *Frontiers in Physiology*, *12*, 712628. https://doi.org/10.3389/fphys.2021.712628

Xu, S., Martinez, L. R., Van Hoof, H., Eljuri, M. I., & Arciniegas, L. (2016). Fluctuating emotions: Relating emotional variability and job satisfaction. *Journal of Applied Social Psychology*, *46*(11), 617–626. https://doi.org/10.1111/jasp.12390

Yerkes, R. M., & Dodson, J. D. (1908). The relation of strength of stimulus to rapidity of habit formation. *Journal of Comparative Neurology & Psychology*, *18*, 459–482. https://doi.org/10.1002/cne.920180503

Zheng, J., Lajoie, S., & Li, S. (2023b). Emotions in self-regulated learning: A critical literature review and meta-analysis. *Frontiers in Psychology*, *14*, 1137010. https://doi.org/10.3389/fpsyg.2023.1137010

Zheng, J., Li, S., & Lajoie, S. P. (2023a). A review of measurements and techniques to study emotion dynamics in learning. In V. Kovanovic, R. Azevedo, D. C. Gibson, & D. lfenthaler (Eds.), *Unobtrusive observations of learning in digital environments* (pp. 7–29). Springer. https://doi.org/10.1007/978-3-031-30992-2_2

Zafeiriou, S., Zhang, C., & Zhang, Z. (2015). A survey on face detection in the wild: Past, present and future. *Computer Vision and Image Understanding*, *138*, 1–24. https://doi.org/10.1016/j.cviu.2015.03.015