Collegiate Aviation Review International

Volume 40 | Issue 2
Fall 2022

The Peer Reviewed Journalof the University Aviation Association

ISSN: 1523-5955

COLLEGIATE AVIATION REVIEW INTERNATIONAL

A PEER REVIEWED JOURNAL OF THE

UNIVERSITY AVIATION ASSOCIATION

COLLEGIATE AVIATION REVIEW INTERNATIONAL
2022 VOLUME 40 ISSUE 2
Yi Gao, Editor

ISSN: 1523-5955

Correspondence and inquiries:

University Aviation Association
8092 Memphis Ave
Millington, TN 38053
(901) 563-0505
hello@uaa.aero

OBJECTIVES

The University Aviation Association publishes the Collegiate Aviation Review International throughout each calendar year. Papers published in each volume and issue are selected from submissions that were subjected to a double-blind peer review process.

The University Aviation Association is the only professional organization representing all levels of the non-engineering/technology element in collegiate aviation education and research. Working through its officers, trustees, committees, and professional staff, the University Aviation Association plays a vital role in collegiate aviation and in the aerospace industry. The University Aviation Association accomplishes its goals through a number of objectives:

- To encourage and promote the attainment of the highest standards in aviation education at the college level
- To provide a means of developing a cadre of aviation experts who make themselves available for such activities as consultation, aviation program evaluation, speaking assignment, and other professional contributions that stimulate and develop aviation education
- To furnish an international vehicle for the dissemination of knowledge relative to aviation among institutions of higher learning and governmental and industrial organizations in the aviation/aerospace field
- To foster the interchange of information among institutions that offer non-engineering oriented aviation programs including business technology, transportation, and education
- To actively support aviation/aerospace oriented teacher education with particular emphasis on the presentation of educational workshops and the development of educational materials covering all disciplines within the aviation and aerospace field

REVIEWER ACKNOWLEDGEMENT

Thank you to all of the individuals who served as reviewers for the scholarly works published in this issue!

Evert de Boer

On Point Loyalty

Patti Clark

Embry-Riddle Aeronautical University

Andrew Dattel

Embry-Riddle Aeronautical University

Derren Duburguet

Metropolitan State University of Denver

Matthew Ebbatson

Swinburne University of Technology *(Australia)*

Paolo Gagliardi

Tauw Italia

Jacqueline Huynh

University of California, Irvine

Jingmin Jin

Aircraft Owners and Pilots Association

Julius Keller

Purdue University

Steven Leib

Central Queensland University (Australia)

Jin-Long Lu

National Kaohsiung University of Science and Technology (Taiwan)

Rian Mehta

Florida Institute of Technology

Flavio Mendonca

Embry-Riddle Aeronautical University

Nikunj Oza

National Aeronautics and Space Administration (NASA)

Matthew Romero

Southern Illinois University - Carbondale

Yi Shao

Oklahoma City University

Sen Wang

Purdue University

Scott Winter

Embry-Riddle Aeronautical University

TABLE OF CONTENTS

Peer-Reviewed Articles

Proceedings of the 2022 UAA Annual Conference

Collegiate Aviation Review International

Volume 40 | Issue 2 Peer Reviewed Article #1

9-21-2022

How Education on Climate Change Affects Consumers' Willingness to Participate in Carbon Offsetting Programs?

Deanna DeMattio
Embry Riddle Aeronautical University

Leila Halawi
Embry Riddle Aeronautical University

If airlines could have a better understanding of how to induce sustainable behavior in their customers, they can use these methods when building sustainable initiatives such as voluntary carbon offsetting programs. The purpose of the study was to determine if education on the consequences of climate change affects how consumers behave related to their interaction with carbon offsetting programs. Regional Scenarios were introduced with a natural disaster comparable to the region under review. A multistage sampling technique was utilized in this study. Firstly, using a cluster sample, groups were used rather than individual units of the target population. The results indicate that education significantly influences behavior compared to behavior before education. Based on the results of this study, we recommend that airlines modify their sustainability initiatives by educating the consumer on the purpose behind the initiative before asking for participation.

Recommended Citation:
DeMattio, D. & Halawi, L. (2022). How education on climate change affects consumers' willingness to participate in carbon offsetting programs? *Collegiate Aviation Review International*, 40(2), 1-18. Retrieved from http://ojs.library.okstate.edu/osu/index.php/CARI/article/view/8664/8407

Introduction

Today, aviation accounts for nearly 3.5% of global carbon dioxide (CO_2) emissions (DLGR, 2021). Companies and regulators within the aviation industry have been working to become more sustainable through nationwide political measures, including the Carbon Offsetting and Reduction Scheme for International Aviation (CORSIA) and the European emissions trading system (ETS) (Wild et al., 2021). Flight taxes and compensation programs are a few political measures created for airlines (Wild et al., 2021). For example, as of June 2021, Switzerland has created a flight tax per traveling passenger from 30 to 120 Swiss Frances (Wild et al., 2021). "Flight tax" is the ticket tax charged at the destination point and is supposed to offer flexibility in the aviation industry (Wild et al., 2021). Although the current study is based in the United States (U.S.), by introducing one of the biggest global carbon markets (ETS), the cap-and-trade system is vital for understanding flight tax (Wild et al., 2021).

The European Economic Area is currently the only market under the control of the EU ETS; however, starting at the end of 2023, the system will apply to all carriers entering the country, including the U.S. (Wild et al., 2021). This system has a fixed cap, which decreases annually. A polluter such as an airline has specific allowances they can emit, for example, two tons of CO_2 (Wild et al., 2021). The allowances can be traded with other airlines to compensate for their emissions. If the company runs out of allowances, a fee is imposed on them, with the goal being for these companies to start investing in clean, renewable energy (Wild et al., 2021).

Carbon offsetting is another market-based approach to reducing CO_2, which refers to lowering emissions of CO_2 to compensate for emissions generated in other parts of the operation (Wild et al., 2021). ICAO adopted CORSIA in 2016 to create this offsetting carbon system. Compared to the EU ETS, CORSIA is the only international aviation climate protection instrument (ICC, 2019). The aviation industry is currently in the pilot phase of CORSIA (2021 - 2023). We will transition into the first phase in 2024, where member states can voluntarily participate in CORSIA (Wild et al., 2021).

Purpose

The study examines whether education on climate change influences consumers' willingness to participate in airlines' carbon offsetting programs.

Importance of the Study

According to Masud et al. (2016), education and awareness determine positive, sustainable behavior; therefore, this study hopes to find a stronger percentage of participants willing to become involved with carbon offsetting programs after education. The airlines within the U.S. would value the information received in this study because they can begin these sustainability programs by evaluating their consumer's knowledge level on climate change and proceed from there to have the most beneficial impact.

Research Question

How does education on climate change influence an individual's willingness to participate in voluntary carbon offsetting programs?

Familiarity with Climate Change

The Intergovernmental Panel on Climate Change (IPCC) has stated that the changing climate problem is one of the more pressing challenges this century (IPCC, 2013). The harm caused by global warming is likely to be unalterable if no action is taken now and into the future to steady the temperature of the Earth's surface. The IPCC report and many other universal environmental organizations believe climate change is artificial. Understanding people's perceptions of climate change is a starting point for this study. However, research has argued that knowledge about climate change is not an essential predictor of concern (Kellstedt et al. 2008).

In contrast, others have determined that knowledge is fundamental (Kellstedt et al., 2008) (Eden, 1998). A 2015 study from the Institute for Environmental Decisions in Zurich, Switzerland, examined how important comprehension of climate change is for people's concern about climate change. The finding was comparable with previous studies showing that scientific knowledge may influence public concern and attitude about climate change (Shi et al., 2015). Shi et al. (2015) research also suggested that people who believe they cannot do anything about a hazard will not change their behavior even when they have more knowledge. Airline consumers knowledgeable about climate change might not be willing to give up their frequent flyer miles until they understand that their contribution would significantly help airlines fund carbon offsetting programs. To change people's behaviors regarding climate change, Shi et al.'s (2015) research suggests that people should be provided with casual knowledge (i.e., information on CO2, temperature warming trends, and climate changes).

Voluntary Carbon Offsetting Programs

Voluntary carbon offsetting programs (VCOP) give eco-minded people the opportunity to compensate for their carbon emissions emitted while flying (Kerner and Brudermann, 2021). Due to increased awareness of climate change, the expectation is to see a higher VCOP. The main idea behind VCOP is to compensate for as many carbon emissions caused by the initial action. This carbon-neutral idea would allow passengers to support airline programs that ensure they take the initiative to offset the emitted carbon of that person's flight. As of today, purchases of VCO have remained low. Studies estimate that less than 10% of passengers have offset at least one flight, and the percent of flights that have been carbon neutral is in the single-digit range (Mair, 2011). This could be from a lack of education about VCOP or an opportunity to participate in the airline industry. Lu and Wang (2018) assume that the sparse carbon offsetting purchases is related to a lack of understanding of the programs and awareness of climate change implications. Current VCOP has no obligation for a consumer, but theoretically, most people should be interested in offsetting their emissions, as climate change will affect all humans in one way (Lu and Wang, 2018). VCOP has been a way to reduce "flight shaming," which is the unease with using air travel for fear of climate implications and public opinion. Purchasers of VCOP have used these programs to travel to distant locations by airplane while minimizing personal carbon emissions.

Mair (2011) created a study to determine what type of people are purchasing VCOP and if those purchasing these programs also show signs of pro-environmental behavior in other ways. The study suggests that individuals who are already willing to contribute to sustainable initiatives are the best audience to target. In a 2007 study, Metz et al. (2007) stated that VCOP needs to be relatively low cost and easy to access to positively attract individuals. A VCOP suggested in this study enables passengers to donate back their frequent flyer miles and loyalty program points to the airline to sustain VCOP. Today, some airlines offer the option to purchase carbon offset credits in ticket transactions (e.g., Qantas and British Airways) (Mair, 2011). Examples of carbon offset programs discussed in Mair's study were avoiding deforestation, afforestation and reforestation, fugitive gases, fossil fuel substitution, and energy efficiency (Mair, 2011). Mair's (2011) exploring air travelers' VCO behavior tested if respondents had offset a flight by purchasing a COP. The results indicated that only 10% of the respondents had purchased, 80% had not purchased VCO before, and 10% were unsure (Mair, 2011). Unfortunately, this and similar studies demonstrate that passengers do not show much willingness to adopt voluntary mitigation initiatives. In contrast, some studies show passengers may only be WTP or act if other passengers do (Winter et al., 2021) (McKercher et al., 2010) (Brouwer et al., 2008). This further exemplifies understanding if education would make a difference in passengers' WTP and act in sustainable initiatives.

Scenario Background

Masud et al. (2016), Pearson and Hamilton (2014), and Rasoli et al. (2019) indicated that introducing a scenario and retesting the theory of planned behavior (TPB) is a way to compare behavioral intention with the same participant. Educational material was collected from the Third National Climate Assessment, which breaks down climate influences in the U.S. region (Horton et al., 2014). This assessment categorized each region of the U.S. to determine three to five notable natural devastating events caused by climate change currently impacting that region. The assessment expands on what will happen to this region if no action exists to extenuate climate change. The following are excerpts from the scenarios and continued literature from the assessment

The introduction of the scenario section included the basic science of climate change, stating that the known cause has been arising in burning fossil fuels since the industrial revolution (UNFCCC, 2015). Discuss further the concern that if the global temperature increases above 2C, natural disasters are expected to become stronger and more frequent (UNFCCC, 2015). Despite the airline's efforts to pursue sustainable initiatives, aviation accounts for around 3.5% of global CO_2 emissions (DLR, 2021). VCOP could help mitigate this amount of CO_2 by taking CO_2 from the atmosphere and investing in sustainable aviation fuels and projects (Kerner and Brudermann, 2021).

The northeastern region of the U.S. not only holds some of the world's more developed cities but is home to major financial centers, the nation's capital, and multiple defining historical landmarks (Horton et al., 2014). One natural disaster that impacted the N.E. was Hurricane Irene, causing massive coastal damage, storm surge, and flooding along the N.E. coastline (Horton et al., 2014). During its impact, Irene produced two to three inches of rainfall per hour in certain areas in late August 2011. A natural disaster, Hurricane Sandy, followed in October 2012, becoming the second most expensive Atlantic hurricane in history (Horton et al., 2014). The northeastern hurricanes, extreme winters, record-breaking heat waves, and flooding cities are due to changes in the Earth's temperature from trapped CO_2 in our atmosphere (Horton et al., 2014).

The midwestern region is home to more than 61 million people with expansive agricultural lands, forests, lakes, and populous cities (Pryor et al., 2014). in the upcoming years, the concern with living in the M.W. has stronger heat wave intensity and frequency, leading to the intense humidity, worsened air quality, and lower water quality affecting public health (Pryor et al., 2014). Prior studies indicate that there will be an increase of up to 2,217 excess deaths per year due to heat stress in Chicago alone by 2081 (Pryor et al., 2014). The M.W. is also characterized by a rich and diverse forest full of natural ecosystems, wetlands, and native species (Pryor et al., 2014). M.W. forests are an amazing absorber of CO_2 and absorb more CO_2 than they emit (Pryor et al., 2014). With weather patterns causing record-breaking high temperatures, insect outbreaks, and increased humidity, this area might soon be changed from a carbon absorber to a carbon-emitting region (Pryor et al., 2014).

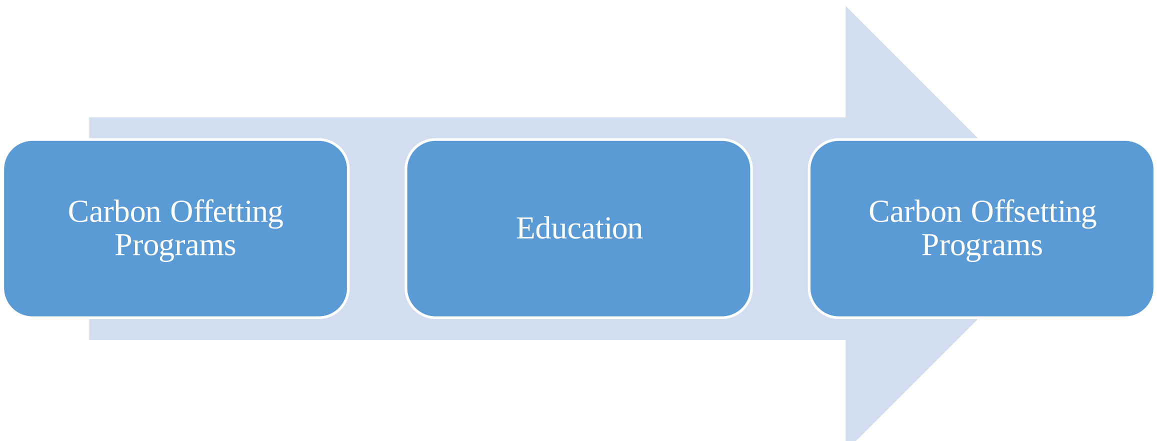
The southern region of the U.S. draws in millions of visitors each year (Carter et al., 2014). One primary concern is reduced water availability due to increased evaporation from increasing temperatures in the upcoming years. With the projected increase in population in the S.E., the continued expansion of the urbanized area will expand these residents' water needs and potentially threaten freshwater qualifiers by exacerbating saltwater during urbanization. Higher sea levels speed saltwater intrusion into freshwater supplies like rivers, streams, and wells. City officials in Hallandale Beach, Florida, have already announced the abandonment of six of their eight drinking water wells (Berry et al., 2011). Food security and increased demand for water are directly related to changes within our southeastern ecosystem. Temperature increase caused by carbon emissions has already shown signs of a War on food and water in the S.E. (Berry et al., 2011).

The southwestern region of the U.S. is the hottest and driest region, while the northwestern region has a complex climate, including snow, fires, and volcanic activity. Increased wildfire risk will alter N.W. forests and increase tree disease outbreaks caused by climate change (Mote et al., 2014). Relative to the 1916 to 2007 period, by the 2080s, the median annual area burned is expected to quadruple to two million acres (Mote et al., 2014). To compensate for the high temperatures, extensive air conditioning can quickly increase the electricity demand and lead to energy system failure or blackouts where the city shuts down the power grid (Mote et al., 2014). Heat stress is already a reoccurring problem for urban residents (Mote et al., 2014). As temperatures are expected to increase, if nothing is done to improve the electricity grid, rolling blackouts could be expected to occur much longer (Mote et al., 2014). Wildfires brought on by these excessive high-temperature days create human respiratory issues, reduce air quality from chemical reactions occurring faster, and cause more disease (Garfin et al., 2014).

Methodology

Using the model from Masud et al. (2016), Pearson and Hamilton (2014), and Rasoli et al. (2019), scenarios explaining climate change disasters in different regions of the country were introduced. The purpose was to determine if participants behaved positively after education on climate change. The fully hypothesized structural model of the factors of carbon offsetting programs before and after education is presented in Figure 1. The arrows indicate a proposed relationship between constructs.

Figure 1
The Structural Model



The use of scenarios in this study is factual rather than a fiction story, like in Rasoli's (2019) study. All regional scenarios introduced a natural disaster comparable to each region. Refer to Appendix A for the study survey. The scenario inclusion was adapted from Masud et al. (2016), Pearson and Hamilton (2014), and Rasoli et al. (2019). The participants completed the nine demographic questions, followed by two questions, one education scenario, and then repeated the survey. Information from each scenario came from (Horton et al.), (Pryor et al., 2014), (Carter et al., 2014), (Lincoln et al., 2021), (Mote et al., 2014), and (Garfin et al., 2014). Each scenario went through a natural disaster caused by climate change in that region. Each region's information was summarized and given to the participants. Participants were only required to read the regional scenario in which they reside. However, they were encouraged to learn about the different natural disasters impacting the U.S. Scenarios were broken down into the northeast (N.E.) region, Southeast (S.E.) region, Midwest (M.W.) Region, and West region. The NE region consisted of M.D., DE, NJ, CT, RI, MA, VT, PA, NY, NH, and M.E. The S.E. region consisted of A, WV, KY, TN, GA, NC, SC, FL, AL, MS, LA, AR, OK, and TX. The M.W. region consisted of N.D., SD, NE, KS, MN, IA, MO, WI, IL, IN, MI, and O.H. The West region consisted of A.K., HI, WA, OR, MT, ID, WY, CO, UT, NV, CA, AZ, and N.M. Each scenario explained that region's natural disaster and the devastation it has caused as well as solutions. The Frequent Flyer Program questions in the demographics section were the control variable. Data was collected to see if participants were active in airline or other loyalty programs. It was not required to already participate in a frequent flyer or loyalty program. The VCOP construct combined two survey questions from Mairs's (2011) research exploring air travelers' VCO behavior and Terblanche's (2015) study understanding the perceived benefits of loyalty programs.

The current study compared the means of carbon offsetting behavior from before and after education using a paired t-test. Descriptive statistics were used to report on the returned demographic information.

Participants

Participants were recruited via a link sent out through Instagram and Facebook, by the Women in Aviation Association via email, and through Embry Riddle Aeronautical University organizations through word of mouth. The survey was conducted in the United States in March 2022. The survey was made available to all U.S. residents to achieve the most comprehensive results for airlines around the country to use. Two hundred thirty-seven responded to the questionnaire, and 97 completed both surveys.

Results

Respondent Demographics

The sample consisted of 57.73% males (n =56), 42.26 % females (n = 41), .87% nonbinary (n = 1), and .87% (n = 1), people who did not report their gender. Concerning marital status, 29.6% of the respondents were married**,** and 60.9% of the participants were single. The reported education level was less than high school (0%, N= 0), high school graduate (3.6%, N= 4), some college (no degree) (10.4%, N= 12), nonformal education (1%, N=.87), associate degree (9.6%, N= 11), bachelor's degree (52.2%, N= 60), master's degree (15.7%, N= 18), doctorate (7%, N= 8). The education level in our sample was higher than that of the general population (Educational Attainment, 2017).

Participants ranged from Boomers to Generation Z. Most of the responses corresponded are Millennials (45.21%), followed by Generation Z with 26%. Participants' primary travel purpose was for leisure/vacation. Participants primarily traveled individually. Participants mean traveling time is twice a year. Participants ranged from the usage of frequent flyer programs to other loyalty programs. 45.2% of participants use frequent airline flyer programs.

Participants were residents from all regions of the U.S. The distribution of respondents by region is reflected in Table 1.

Table 1
Respondents by Region

Region	Frequency	Percent of Total	Cumulative Percent of Total
Midwest	9	8.2568807	8.2568807
Northeast	28	25.688073	33.944954
South	35	32.110092	66.055046
West	36	33.027523	99.082569

Hypothesis 1

This research seeks to test the following hypothesis：

H1a: There is a difference in behavior to volunteer for carbon offset programs between individuals educated about climate change and those not educated.

H10: There is no difference in behavior to volunteer for carbon offset programs between individuals educated about climate change and those not educated.

A paired t-test (Table 2) was used to show whether behavior's sample means after education is different than the sample means of behavior before education. A t-test for dependent means was used because the research constructs were repeated. The comparison distribution is a t-distribution with N-1 degrees of freedom = 96 *df*. This is a paired test with 96 degrees of freedom, and we tested at the .05 significance level. We received a p-value of 0.002. We reject the null hypothesis. There is a significant difference in behavior to volunteer for carbon offset programs between individuals educated about climate change and those not educated. Hypothesis 1 is supported.

Table 2
Paired T-Test Results

$\mu D = \mu 1 - \mu 2$: Mean of the difference between COP 1 and COP 2
$H0 : \mu D = 0$
$HA: \mu D \neq 0$

Sample Statistics

Sample	n	Mean	Std. Dev.
COP 1	97	2.9347079	1.2106197
COP 2	97	2.7800687	1.3027185

Hypothesis Test Results:

Difference	Mean	Std. Err.	DF	T-Stat	P-value
COP 1 - COP 2	0.15463918	0.048621179	96	3.18049	0.002

Limitations

This analysis is not without any limitations. This study represents data from U.S residents over 18. This consideration may limit the generalizability of the findings beyond this specific sample. The analysis was also based on preference data that may present hidden biases that highlight instinctual ways individuals make selections. Even though this work examines whether education on climate change influences consumers' willingness to participate in airlines' carbon offsetting programs, we cannot confirm that this translates into actual behavior change.

Conclusion

A comprehensive literature review was completed, factors contributing to climate change initiatives were identified, and one hypothesis was tested. A survey instrument was used to collect data from a sample of travelers within the U.S. Our hypothesis was supported through a t-test and contained specific behaviors like, "I would purchase a carbon offset for a flight." Participants understood the behavior and acted based on their newfound education and knowledge rather than their emotions, ethics, or morals.

Recommendations for Practice

Airlines are constantly working towards the 2050 goal of carbon-neutral within the aviation industry. One key component to carbon neutrality is using good, reliable carbon offsetting programs. Knowing that airlines can receive assistance from their consumers through education in funding these projects and many others is valuable to the industry. Understanding what affects their behavior is an important and challenging component of getting stakeholder and consumer engagement. This study can help airlines understand the importance of education and its impact on behavior change of their consumers and stakeholders.

Recommendations for Future Research

Multiple paths for future research have surfaced since the beginning of our study and results from the data. First, we collected a convenience sample of data from only U.S. residents. We suggest a targeted sample be used to find individuals who travel most often or use reward programs frequently. To advise airlines, it would be best to gather data from people who support the suggested behavior the best, giving back points/miles to support initiatives.

Second, , a different source of education that is still related to climate change could be used to test the model. This study focused on natural disasters in the resident's area, whereas other research could educate the applicant about changes to agriculture or transportation from ongoing climate change.

References

Anderson, G. B., and M. L. Bell, 2011: Heatwaves Heatwaves in the United States: Mortality risk during heat waves and effect modification 56. by heat wave characteristics in 43 U.S. communities. *Environmental Health Perspectives*, 119, 210-218, doi:10.1289/ehp.1002313.

Berry, L., F. Bloetscher, N. Hernández Hammer, M. Koch-Rose, D. Mitsova-Boneva, J. Restrepo, T. Root, and R. Teegavarapu, 2011: Florida Water Management and Adaptation in the Face of Climate Change, 68 pp., Florida Climate Change Task Force. Available online at http://floridaclimate.org/docs/water_managment.pdf

Brouwer, R., Brander, L., & Van Beukering, P. (2008). "A convenient truth": Air travel passengers' willingness to pay to offset their CO2 emissions. *Climatic Change, 90*, 299–313.

Busch, T., & Judick, L. (2021). Climate change—that is not real! A comparative analysis of Climate-skeptics-sceptic think tanks in the USA and Germany. *Climatic Change, 164*(1-2) https://doi.org/10.1007/s10584-021-02962-z

Carter, L. M., J. W. Jones, L. Berry, V. Burkett, J. F. Murley, J. Obeysekera, P. J. Schramm, and D. Wear, 2014: Ch. 17: Southeast and the Caribbean. *Climate Change Impacts in the United States: The Third National Climate Assessment*, J. M. Melillo, Terese (T.C.) Richmond, and G. W. Yohe, Eds., U.S. Global Change Research Program, 396-417. doi:10.7930/J0N- P22CB.

DGLR, 2021. Sustainability in Aviation. In: Rötger, T. (Ed.), *Aviation & Aerospace*, DGLR edition 1. DGLR, Bonn(in German), 29–33.

Eden S. Environmental issues: Knowledge, uncertainty and the environment. Progress in Human Geography, 1998; 22(3): 425–432.

Educational Attainment, by Race and Ethnicity - Race and Ethnicity in Higher Education. Race and Ethnicity in Higher Education. (2017). https://www.equityinhighered.org/indicators/u-s-population-trends-and-educational-attainment/educational-attainment-by-race-and-ethnicity/

Garfin, G., G. Franco, H. Blanco, A. Comrie, P. Gonzalez, T. Piechota, R. Smyth, and R. Waskom, 2014: Ch. 20: Southwest. *Climate Change Impacts in the United States: The Third National Climate Assessment*, J. M. Melillo, Terese (T.C.) Richmond, and G. W. Yohe, Eds., U.S. Global Change Research Program, 462-486. doi:10.7930/J08G8HMN.

Horton, R., G. Yohe, W. Easterling, R. Kates, M. Ruth, E. Sussman, A. Whelchel, D. Wolfe, and F. Lipschultz, 2014: Ch. 16: Northeast. *Climate Change Impacts in the United States: The Third National Climate Assessment,* J. M. Melillo, Terese (T.C.) Richmond, and G. W. Yohe, Eds., U.S. Global Change Research Program, 16-1-nn.

IATA. Airlines expect a 31% rise in passenger demand by 2017. Available online: http://www.iata.org/pressroom/pr/pages/2013-12-10-01.aspx

ICC, 2019. ICC signs a partnership at COP25 to support a carbon-neutral aviation industry. Available online: https://iccwbo.org/media-wall/news-speeches/icc-signs-partners hip-at-cop25-to-support-a-carbon-neutral-aviation-industry

International Chamber of Commerce, ICC, Headquarters, Paris, Fr.

IPCC, 2013: Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge, UK, and New York, NY: Cambridge University Press 2013.

Kellstedt PM, Zahran S, Vedlitz A. Personal efficacy, the information environment, and attitudes toward global warming and climate change in the United States. Risk Analysis, 2008; 28(1):113–126.

Kerner, C.; Brudermann. I Believe I Can Fly—Conceptual Foundations for Behavioral Rebound Effects Related to Voluntary Carbon Offsetting of Air Travel. *Sustainability* 2021,*13*,4774. https://doi.org/ 10.3390/su13094774

Lincoln Lenderking, H., Robinson, S., & Carlson, G. (2021). Climate change and food security in caribbeanCaribbean small island developing states Challenges and strategies. *International Journal of Sustainable Development and World Ecology, 28*(3), 238-245. https://doi.org/10.1080/13504509.2020.1804477

Lu, J., & Wang, C. (2018). Investigating the impacts of air travelers' environmental knowledge on attitudes toward carbon offsetting and willingness to mitigate the environmental impacts of aviation. *Transportation Research. Part D, Transport and Environment, 59,* 96-107. https://doi.org/10.1016/j.trd.2017.12.024

Mair, J. (2011) Exploring air travelers' voluntary carbon-offsetting behavior, Journal of Sustainable Tourism, 19:2, 215-230

Masud, M. M., Al-Amin, A. Q., Junsheng, H., Ahmed, F., Yahaya, S. R., Akhtar, R., & Banna, H. (2016). Climate change issue and theory of planned behavior: Relationship by empirical evidence. *Journal of Cleaner Production, 113,* 613-623.

McKercher, B., Prideaux, B., Cheung, C., & Law, R. (2010). Achieving voluntary reductions in the carbon footprint of tourism and climate change. *Journal of Sustainable Tourism, 18*(3), 297–317.

Metz, B., Davidson, O.R., Bosch, P.R., Dave, R., & Meyer, R.A. (Eds.). (2007). Climate change 2007: *Mitigation. Contribution of working group III to the fourth assessment report of the IPCC. Cambridge: Cambridge University Press; IPCC.*

Mote, P., A. K. Snover, S. Capalbo, S. D. Eigenbrode, P. Glick, J. Littell, R. Raymondi, and S. Reeder, 2014: Ch. 21: North- west. *Climate Change Impacts in the United States: The Third National Climate Assessment*, J. M. Melillo, Terese (T.C.) Rich- mond, and G. W. Yohe, Eds., U.S. Global Change Research Program, 487-513. doi:10.7930/J04Q7RWX.

Nunez, S., Arets, E., Alkemade, R., Verwer, C., & Leemans, R. (2019). Assessing the impacts of climate change on biodiversity: Is below 2 °C enough? *Climatic Change*, *154*(3-4), 351-365. https://doi.org/10.1007/s10584-019-02420-x

Pearson, M., & Hamilton, K. (2014). Investigating driver willingness to drive through flooded waterways. *Accident Analysis and Prevention, 72*, 382-390. https://doi.org/10.1016/j.aap.2014.07.018

Pryor, S. C., D. Scavia, C. Downer, M. Gaden, L. Iverson, R. Nordstrom, J. Patz, and G. P. Robertson, 2014: Ch. 18: Midwest. *Climate Change Impacts in the United States: The Third National Climate Assessment*, J. M. Melillo, Terese (T.C.) Rich- mond, and G. W. Yohe, Eds., U.S. Global Change Research Program, 418-440. doi:10.7930/J0J1012N.

Rasoli, M., Mirrezaie, S. M., Fooladi, E., Hosseini, R. Z., & Fayaz, M. (2019). Effects of reviewing childbirth scenarios on the choice of delivery type: A randomized controlled trial. *Turkish Journal of Obstetrics and Gynecology*, *16*(1), 15-22.

Shi, J., Visschers, V. H. M., & Siegrist, M. (2015). Public perception of climate change: The importance of knowledge and cultural worldviews. *Risk Analysis*, *35*(12), 2183-2201. https://doi.org/10.1111/risa.12406

UNFCCC (2015) Adoption of the Paris Agreement. Report No. FCCC/CP/2015/L.9/Rev.1. UNFCCC, Bonn, http://unfccc.int/resource/docs/2015/cop21/eng/l09r01.pdf

Wild, P., Mathys, F., & Wang, J. (2021). Impact of political and market-based measures on aviation emissions and passenger behaviors (a swiss case study). *Transportation Research Interdisciplinary Perspectives*, *10*, 100405.

Winter, S., Crouse, S., & Rice, S. (2021). *The development of 'green' airports: Which factors influence willingness to pay for sustainability and intention to act? A structural and mediation model analysis*. Science Direct.

Appendix A Questionnaire

Descriptive and demographic statistics of study participants

Gender
- Male
- Female
- Nonbinary
- Transmasculine
- Other

Ethnicity
- Caucasian
- African Descent
- Hispanic Descent
- Asian Descent
- Other
- No Response

Education
- Less than high school
- High School graduate
- Some college (no degree)
- Non-Formal Education
- Associate Degree
- Bachelor's Degree
- Master's Degree
- Doctorate Degree

Age (Born in)
- After 2012
- 1997 – 2012
- 1981 -1996
- 1965 -1980
- 1955 -1964
- 1946 -1954

Marital Status
- Single
- Married
- Divorced
- Separated
- Widow or widower
- Other

Primary Travel Purpose
- Leisure/ Vacation
- Business
- Conference/training
- Visiting family/friends
- Academic/study
- Other

Primary Travel
- Individual
- In a group
- No Response

Frequency of Travel
- 1 – 3 Times a Week
- More than 3 Times a Week
- 1-3 Times a Month
- More than 3 Times a Month
- Once a year
- Twice a year
- Never
- No response

Frequent Flyer Program
(Ref. (Terblanche, 2015)
- I use airline frequent flyer programs
- I don't use airline frequent flyer programs
- I use other company loyalty programs
- I don't use other company loyalty programs

Voluntary Carbon Offsetting Program (VCOP) Ref. (Terblanche, 2015 and Mair, 2011)	COP1: I would purchase a carbon offset for a flight.	1	2	3	4	5
	COP2: I would give up my loyalty program points from any company to offset my travel emissions.	1	2	3	4	5

Scenarios

Introduction

We are going to kick this off with some basic science. Climate change is widely due to the increase in burning fossils fuels since the industrial revolution. Our population has also tripled in the past 70 years (UNFCCC, 2015), which has caused for larger production of these fossil fuels. These gases are released into the air, and as sunlight hits Earth's surface, some of the heat gets trapped, and the climate gets warmer, also known as the greenhouse effect. The average global temperature has increased by almost 1C since 1880 (Nunez et al., 2019). The 2015 Paris Agreement set forth a goal to keep global temperature increase below 2C (UNFCCC, 2015). Changes in the global temperature have led to regional changes in our climate and extreme natural disasters such as drought, flooding, wildfires, water shortages, and rising sea levels from glaciers melting. People are losing their homes, cities are flooding, islands are disappearing, and the worst is yet to occur (Busch and Judick, 2021). Scientists have referred to this catastrophe as the 6th mass extinction event.

Presently, aviation accounts for around 3.5% of global CO2 emissions (DLGR, 2021). Despite the airline's efforts in pursuing sustainable initiatives, the responsibility of the airline industry continues to grow due to rapid growth in airline travel (IATA, 2017). Voluntary carbon offsetting programs VCOP give eco-minded people the opportunity to compensate for their carbon emissions emitted while flying (Kerner and Brudermann, 2021). A carbon offset is an investment in a climate change mitigating project. The main idea behind VCOP is to compensate for as many carbon emissions caused by the initial action. This carbon-neutral idea would allow passengers to support airline programs that ensure they take the initiative to offset the emitted carbon of that person's flight. The offsetting carbon program we propose is giving back your frequent flyer miles to support many of the sustainable initiatives that airlines are set to achieve with a guaranteed promise your loyalty program rewards will be used for said project.

Please choose the region in which you reside to learn more about the changes in climate in your area. Every region is experiencing similar but different natural disasters, and we encourage you to learn about them all. Please feel free to read about all four regions but only select the one region you reside in.

Northeast: M.D., DE, NJ, CT, RI, MA, VT, PA, NY, NH, ME

Heat waves, sea-level rise, and cold air outbreaks pose a growing risk to the cities and towns in the Northeast (N.E.) (Horton et al., 2014). Sixty-four million people are living in the N.E. today. This area not only holds some of the most developed cities in the world but is home to leading financial centers, the nation's capital, and many defining historical landmarks. This area includes large cities and more than 180,000 farms that bring in $17 billion in annual sales (Horton et al., 2014).

One natural disaster that impacted the N.E. was Hurricane Irene, causing massive coastal damage, storm surge, and flooding along the N.E. coastline (Horton et al.). During its impact, Irene produced two to three inches of rainfall per hour in certain areas in late August 2011. Natural disaster, Hurricane Sandy, followed in October 2012, becoming the second most costly Atlantic hurricane in history (Horton et al., 2012). In preparation for Hurricane Irene, New York City transportation was shut down, and over 2.3 million coastal residents were forced to evacuate (Horton et al., 2014). However, the most severe impacts were on the inland areas of upstate New York and central Vermont. Flash flooding wiped out roads and bridges, undermined railroads, brought down trees, powerlines, and buildings, flooded homes, and devastated ecosystems (Horton et al., 2014). In Vermont, over 500 miles of roadways were damaged, estimating \$175 to \$250 million in rebuilding costs (Horton et al., 2014). Residents suffered from mold growth in flooded homes and cleaning up spills from damaged waste tanks cost Vermont an estimated \$1.75 million (Horton et al., 2014). Hurricane Irene took 41 lives, while Hurricane Sandy was responsible for 150 deaths. Sandy cost the Northeast between \$60 to \$80 billion in repairs. 8.5 million people were without power, and an estimated 650,000 homes were damaged after Hurricane Sandy (Horton et al., 2014). Projections and suspected vulnerabilities due to coastal flooding and sea levels rising have been available as early as 2001, and although these reports were around, the devastation was still a surprise to many residents.

Disruption to services interrupt commerce and threaten public health in safety in the N.E. In New York State alone, sea-level rise is estimated to flood or render unusable 2,112 miles of roads, 3,647 acres of airport facilities, and 539 acres of runways (Horton et al., 2014). The northeastern hurricanes, along with extreme winters, record-breaking heat waves, and flooding cities, are due to changes in the Earth's temperature from trapped CO2 in our atmosphere. Carbon offsetting not only enables customers and airline passengers to invest in sustainability but also has the potential to decrease and prevent further natural disasters from occurring in your backyard.

Midwest: N.D., SD, NE, KS, MN, IA, MO, WI, IL, IN, MI, OH

Home to more than 61 million people, the Midwest (M.W.) has expansive agricultural lands, forests, lakes, and populous cities (Pryor et al., 2014). The relationship between people, ecosystems, and infrastructure is at risk with posing climate change. Increased heat stress, flooding, and drought will be multiplied by changes in pests and diseases from temperatures increasing and climates shifting (Pryor et al., 2014). Competition from native and non-native species, agriculture shocks from extreme natural disasters, and landscape changes, when taken collectively, are projected to alter the socioeconomic patterns and ecosystems of the M.W. The Great Lakes and northern forest are major supplies to fisheries, recreation, tourism, and commerce in the M.W.

The concern with living in the M.W. in the upcoming years is increased heat wave intensity and frequency, leading to increased humidity, degraded air quality, and reduced water quality affecting public health (Pryor et al., 2014). The frequency of major heat waves has increased over the last six decades in the M.W. (Pryor et al., 2014). Within the United States, we see mortality increase by 4% during heatwave days compared with non-heat wave days (Anderson and Bell, 2011). On July 20th, 2011, the majority of the M.W. experienced temperatures over 100F, and for most of that summer, the citizens were under a heat alert (Anderson and Bell, 2011). Studies predict that there will be an increase of up to 2,217 excess deaths per year due to heat stress in Chicago alone by 2081 (Pryor et al., 2014). Air quality in most of the M.W. fails to meet the national ambient air quality standards due to induced emissions and increased pollen season, which is predicted to amplify during higher temperatures (Pryor et al., 2014). Increased temperatures also correlate to increased diseases and higher amounts of disease carriers like insects and rodents (Pryor et al., 2014).

The M.W. is also characterized by a rich and diverse forest full of natural ecosystems, wetlands, and native species (Pryor et al., 2014). Global carbon absorbers, such as these forests, are at risk of not intake as much carbon and store it from disturbances in insect outbreaks, fires, and droughts from heatwaves. M.W. forests are an amazing absorber of carbon dioxide (CO2) and absorb more CO2 than they emit (Pryor et al., 2014). With changes in weather patterns causing record-breaking high temperatures, insect outbreaks, and increased humidity, this area might soon be changed from a carbon absorber to a carbon-emitting region (Pryor et al., 2014). Planting trees is one method of carbon offsetting available for the aviation industry. Taking in carbon and allowing these forests to be under less stress enables all the ecosystems in the M.W. to breathe and work on adapting to higher temperatures. Carbon offsetting not only enables customers and airline passengers to invest in sustainability but also has the potential to decrease and prevent further natural disasters, drought, and starvation from occurring in your backyard.

South: V.A., WV, KY, TN, GA, NC, SC, FL, AL, MS, LA, AR, OK, TX

From the Appalachian Mountains to the coastal plains, this area is home to more than 80 million people and draws in millions of visitors each year (Carter et al., 2014). The Gulf and Atlantic coasts are major producers of seafood, home to seven large ports, and are extremely vulnerable to sea-level rise. The number of category 4 and 5 hurricanes has increased substantially since the early 1980s, and in recent years the World Meteorological Organization has run out of phonetic alphabet names and had to use the Greek alphabet to name the remainder of the hurricanes in those years (Carter et al., 2014). Water resources in the Southeast (S.E.) are abundant and support the various populated cities, rural communities, and unique ecosystems (Carter et al., 2014). Water conflicts due to drought have occurred between states, such as the 2007 drought in Atlanta, Georgia (Carter et al., 2014). One major concern noted in the S.E. is reduced water availability in the upcoming years due to increased evaporation from rising temperatures. With the projected increase in population in the S.E., the continued development of the urbanized area will increase these citizens' water needs and potentially threaten freshwater qualifiers by exacerbating saltwater during urbanization. Higher sea levels will accelerate saltwater intrusion into freshwater supplies like rivers, streams, and wells. City officials in Hallandale Beach, Florida, have already announced the abandonment of six of their eight drinking water wells (Berry et al., 2011). Water demand is already needed to increase from an increase in demand for food and agricultural production due to the population increase in the S.E. With a concern of limited water availability already happening, a plan is needed to protect the citizens of the S.E. from protecting them against their human right to clean drinking water.

Rising air and sea surface temperatures, variability in precipitation patterns, increased storms and hurricanes, and impact on water availability also made food security a major concern for the S.E. (Lincoln et al., 2021). Food security is the state in which people, at all times, have access to sufficient, safe, and nutritious food to meet dietary needs for a healthy life (Lincoln et al., 2021). Declines in fisheries, from impacts on ports in the S.E., are a cause for concern about climate change in this region. Mild water deficits can create reduced growth rate, fertilization issues, and overall reduced fruit yield for many farmers in the S.E. The main threat noticed to meat production is heat stress, which decreases animals' productivity, fertility, ability to gain weight, and increases the risk of diseases (Lincoln et al., 2021). Communities that are dependent on fisheries are at risk of having vulnerable fish begin to reproduce less, due to warming ocean temperatures (Lincoln et al., 2021). Food security and increased demand for water are directly related to changes within our southeastern ecosystem. Temperature increase caused by carbon emissions has already shown signs of a War on food and water in the S.E. Carbon offsetting not only enables customers and airline passengers to invest in sustainability but also has the potential to decrease and prevent further natural disasters, drought, and starvation from occurring in your backyard.

West: A.K., HI, WA, OR, MT, ID, WY, CO, UT, NV, CA, AZ, NM

The southwest S.W. is the hottest and driest region in the United States, while the northwest N.W. has a complex climate including snow-packed mountains, fires, and volcanic activity. Granted their differences, the combined impacts of increasing wildfires in both areas are already causing widespread tree die-off and long-term forest landscape transformations (Mote et al., 2014). In the southwestern region, not only do these wildfires pose a threat to the public, but amplified heat causes disruptions to urban electricity and water supplies (Garfin et al., 2014).

Increased wildfire risk will alter N.W. forests and increase tree disease outbreaks caused by climate change (Mote et al., 2014). Relative to the 1916 to 2007 period, by the 2080's the median annual area burned is expected to quadruple to two million acres (Mote et al., 2014). Pine mortality has already increased from a spike in the mountain pine beetle which increases with warmer temperatures (Mote et al., 2014). Increased wildfires are expected to have health consequences by exacerbating respiratory and cardiovascular illnesses in cities that surround these forests. The N.W. economy can see significant impacts from local timber revenues decreasing (Mote et al., 2014).

The S.W. has the highest percentage of its population living in cities and its urban population rate is 12% greater than the national average (Garfin et al., 2014). These urban infrastructures are vulnerable more than normal because of their interdependence, meaning strains in one system can cause disruptions in another (Garfin et al., 2014). To deal with high temperatures, extensive air conditioning can quickly increase the electricity demand and trigger energy system failure or result in blackouts where the city shuts down the power grid. Heat stress is already a reoccurring problem for urban residents. As temperatures are expected to increase, if nothing is done to improve the electricity grid, rolling blackouts could be expected to occur much longer. Wildfires brought on by these excessive high-temperature days, create human respiratory issues, reduce air quality from chemical reactions occurring faster, and more disease (Garfin et al., 2014). These warm seasons are expected to be longer as years go on in the N.W. and S.W. Without changes to our day-to-day sustainable behaviors, reduction in CO2, and education about climate change, the path that the S.W. and N.W. are on is not easy or safe. Carbon offsetting not only enables customers and airline passengers to invest in sustainability but also has the potential to decrease and prevent further natural disasters, drought, and starvation from occurring in your backyard.

Collegiate Aviation Review International

Volume 40 | Issue 2 Peer Reviewed Article #2

10-9-2022

Validating ADS-B Data for Use in Noise Modeling Applications

Luigi Raphael I. Dy
Purdue University

John H. Mott
Purdue University

Aircraft noise continues to be a major environmental issue impacting airports and their surrounding communities. Beyond being an annoyance, aircraft noise has also been found to have potentially adverse health effects on humans and animals. Thus, international, national, and local regulations have been adopted to quantify, limit, and mitigate aircraft noise. Software developed by the Federal Aviation Administration to estimate the impacts of airport noise relies on operations information that may be difficult to obtain for aircraft operating under visual flight rules at non-towered airports. Hence, leveraging the use of ADS-B as a low-cost source of operations data may improve noise estimation methods at such airports. To validate this approach, ADS-B data was compared to GPS records from aircraft avionics. With an average error of 57.72 feet laterally, 112.36 feet vertically, and 126.32 feet combined, resulting noise estimation errors as a result of ADS-B position errors are expected to be less than seven decibels. It was also found that ADS-B data can be significantly improved by incorporating atmospheric data to improve altitude information, leading to a reduction in estimation errors. The results of this study highlight the potential applicability of ADS-B usage in noise estimation and other applications.

Recommended Citation:
Dy, L. R. I. & Mott, J. H. (2022). Validating ADS-B data for use in noise modeling applications. *Collegiate Aviation Review International*, 40(2), 19-32. Retrieved from http://ojs.library.okstate.edu/osu/index.php/CARI/article/view/8663/8414

Introduction

Noise is a negative byproduct of aircraft operations that affects not only individuals at airports, but also in surrounding communities. In addition to being an annoyance, environmental noise such as that from flying aircraft has been found to have detrimental effects on the health and development of humans and animals. Thus, noise mitigation, abatement procedures, policies, and regulations have been studied and implemented at airports around the world. Aside from sound meters, noise models based on flight track data and aircraft information are used to study aircraft noise. However, non-towered airports, which comprise the largest proportion of U.S. airports, do not have access to these pertinent data due to the lack of radar services and tower facilities. As such, the use of ADS-B data collected from low-cost devices has been proposed as a method to collect information necessary to model noise, enabling non-towered airports to leverage noise modeling systems (Yang & Mott, 2022). Huynh et al. (2022) have similarly used ADS-B data to model aircraft approach performance to evaluate the benefit of delayed deceleration approaches on aircraft noise impacts, while Gagliardi et al. (2017) have used ADS-B data to evaluate the compliance rate and effect of noise abatement procedures.

This study provides a quantification of the errors associated with ADS-B data when compared to recorded onboard position information and was performed to provide a sense of the errors from noise estimation that may be attributable to the use of ADS-B as the data source. A total of 246,349 position reports from a sample of 60 flights conducted by all 13 Piper Archers of the Purdue University fleet over a five-month period were used. Errors were calculated by comparing received ADS-B data with corresponding data collected from onboard flight recorders built into the aircraft's Garmin G1000Nxi avionics package.

Background

In a review conducted by Basner et al. (2014), acute and chronic exposure to noise has been established to cause adverse auditory effects in humans, including hearing loss and symptoms such as tinnitus. In addition, noise exposure has been found to increase the risk of cardiovascular disease, decrease cognitive performance, and negatively affect the quality of sleep. Negative impacts on sleep further affect risks for cardiovascular diseases, diabetes, and weight gain and obesity (Swift, 2010). Meta-analyses have shown a 7% to 17% increase in the risk of cardiovascular conditions and diseases per 10 dB increase in exposure to transportation noise, after adjusting for known risk factors (Basner et al., 2014). Studies have also shown negative effects of transportation noise such as those from air and road traffic on children's learning outcomes, cognitive performance, reading ability, memory, standardized testing performance, and endocrine responses (Basner et al., 2014; Stansfeld & Clark, 2015). Thus, efforts to reduce noise exposure are expected to reduce annoyance, improve children's development, improve sleep, and lower the prevalence of cardiovascular diseases, while improvements in noise modeling and monitoring can only further improve our understanding of noise and its effects (Basner et al., 2014; Stansfeld & Clark, 2015).

In an older study by Manci et al. (1988), noise effects on different domestic and wild animals were established to include fright, hearing loss, decreased fertility, and hypertension. Thus, the effect of aircraft noise on wildlife and the environment must also be considered.

Pertinent Regulations and Standards

Due to the apparent and potential effects of aircraft noise exposure, the United States has federal noise control regulations and guidance in place, identifying public health as the physiological and psychological well-being of the public, the effects of noise and the common interest of the community are to be considered (Penn State, n.d.). Similarly, the International Civil Aviation Organization (ICAO) has adopted Annex 16, Volume I – Aircraft Noise as a standard for aircraft noise regulations (ICAO, n.d.).

To quantify aircraft noise exposure, the Federal Aviation Administration (FAA) uses the Day-Night Average Sound Level (DNL) as a primary metric, in addition to other commonly used metrics such as Sound Exposure Level (SEL), Sound Pressure Level (SPL), Equivalent Sound Level (L_{eq}), Maximum A-weighted Sound Level (L_{max}), Time Above a Specified Sound Level (TA), and Number of Events Above a Specified Noise Level (NA) (Penn State, n.d.). These metrics are calculated through sound level meters deployed around airports, around aircraft flight paths or through models that estimate the amount of noise received at a specific location based on air traffic. Meanwhile, the ICAO and FAA use Effective Perceived Noise Level (EPNL) for the purpose of aircraft noise compliance certification (Appendix A to Noise Standards, 2021; ICAO, n.d.).

With regard to aircraft type certification, Title 14 of the Code of Federal Regulations (14 CFR) Part 36 contains the noise limits to which different categories of aircraft must conform for type certification. In addition, it outlines the procedures and adjustments to be made during the data collection process. (Appendix A to Noise Standards, 2021).

For airports, 14 CFR Part 150 contains regulations pertaining to the development and use of airport noise exposure maps and airport compatibility programs, in addition to information and regulations defining compatible land use around airports as a result of noise created by aircraft operations. Noise measurement standards are outlined in the appendix of the regulation (Appendix A to Airport Noise Compatibility Planning, 2021). For example, appendix A of Part 150 specifies the use and calculation of Yearly Day-Night Average Level (YDNL) and the need for continuous 65, 70, and 75 YDNL contours to be developed for an airport through measurement or estimation. Using these noise contours, the determination as to whether the land use is “compatible” based on Federal regulations as seen in Figure 1, or pertinent local laws, can be made. Appendix A of Part 150 further specifies that to develop a computer-modeled noise exposure map, flight tracks, fleet mix, operational data and trends, and approach paths must be collected, among other parameters, for use with an FAA-approved method or software.

Figure 1

Table 1 of Appendix A of 14 CFR Part 150

TABLE 1—LAND USE COMPATIBILITY* WITH YEARLY DAY-NIGHT AVERAGE SOUND LEVELS

Land use	Yearly day-night average sound level (L_{dn}) in decibels					
	Below 65	65-70	70-75	75-80	80-85	Over 85
RESIDENTIAL						
Residential, other than mobile homes and transient lodgings	Y	N(1)	N(1)	N	N	N
Mobile home parks	Y	N	N	N	N	N
Transient lodgings	Y	N(1)	N(1)	N(1)	N	N
PUBLIC USE						
Schools	Y	N(1)	N(1)	N	N	N
Hospitals and nursing homes	Y	25	30	N	N	N
Churches, auditoriums, and concert halls	Y	25	30	N	N	N
Governmental services	Y	Y	25	30	N	N
Transportation	Y	Y	Y(2)	Y(3)	Y(4)	Y(4)
Parking	Y	Y	Y(2)	Y(3)	Y(4)	N
COMMERCIAL USE						
Offices, business and professional	Y	Y	25	30	N	N
Wholesale and retail—building materials, hardware and farm equipment	Y	Y	Y(2)	Y(3)	Y(4)	N
Retail trade—general	Y	Y	25	30	N	N
Utilities	Y	Y	Y(2)	Y(3)	Y(4)	N
Communication	Y	Y	25	30	N	N
MANUFACTURING AND PRODUCTION						
Manufacturing, general	Y	Y	Y(2)	Y(3)	Y(4)	N
Photographic and optical	Y	Y	25	30	N	N
Agriculture (except livestock) and forestry	Y	Y(6)	Y(7)	Y(8)	Y(8)	Y(8)
Livestock farming and breeding	Y	Y(6)	Y(7)	N	N	N
Mining and fishing, resource production and extraction	Y	Y	Y	Y	Y	Y
RECREATIONAL						
Outdoor sports arenas and spectator sports	Y	Y(5)	Y(5)	N	N	N
Outdoor music shells, amphitheaters	Y	N	N	N	N	N
Nature exhibits and zoos	Y	Y	N	N	N	N
Amusements, parks, resorts and camps	Y	Y	Y	N	N	N
Golf courses, riding stables and water recreation	Y	Y	25	30	N	N

Note. This figure shows the maximum allowable Yearly Day-Night Average Levels (YDNL) for different types of land uses. Y indicates the land use is allowable at that YDNL and N indicates that it is not. Parentheses indicate notes (included in the appendix of the regulation) while the numbers 25, 30, or 35 indicate the noise level attenuation designed or built into the structure to be acceptable. Table adapted from "Appendix A to Airport Noise Compatibility Planning", 14 C.F.R. § A150, 2021 (https://www.ecfr.gov/current/title-14/chapter-I/subchapter-I/part-150).

Current Methods

The primary noise model that the FAA uses is part of the Aviation Environmental Design Tool (AEDT) software available for airport operators and related users that support modeling and estimation of environmental impacts such as fuel consumption, emissions, air quality impacts, and noise of individual flights, or flights on a larger regional or national scale (FAA, n.d.-a). The noise model part of AEDT replaced the Integrated Noise Model (INM) previously used by the FAA and specified in 14 CFR Part 150 (FAA, 2019). AEDT is comprehensive and can be extremely detailed. Information such as the number and type of ground service equipment utilized for a specific operation can be provided to improve estimates of its environmental impacts, while for noise modeling, an acoustic impedance adjustment option is available to correct for noise propagation characteristics due to non-standard meteorological conditions (FAA, 2021a; FAA, 2021b)

Users of AEDT add flights to a study by specifying events and conditions, including the amount of time spent on the ground and taxiing, the flight track, the type of operation, the aircraft and its engines, the takeoff thrust settings used, the percentage of flights that perform touch and goes, and in multi-aircraft studies, the fleet mix (FAA, 2021a). This information is used by AEDT to calculate estimated noise impacts on the surrounding areas. AEDT does this by utilizing information from European Organisation for the Safety of Air Navigation's (EUROCONTROL's) Base of Aircraft Data (BADA) to approximate the actual performance and flight profiles of the aircraft as realistically as possible based on the input data. The approximated flight paths and climb/descent profiles are then used in conjunction with EUROCONTROL's Aircraft Noise and Performance (ANP) database to estimate noise exposure metrics and impacts (FAA, 2021b). These features allow users to produce detailed and accurate environmental impact estimates based on the input data.

While AEDT is an established tool that gives airport operators and stakeholders detailed estimates of the environmental impacts of airport operations, its output is only as good as its input. Thus to be accurate and complete, information is needed from the user. However, for non-towered airports, information such as fleet mix and number of operations are often limited to estimates and approximations as compared to towered airports that have air traffic controllers that tally the number of operations. Additionally, as aircraft flying under Visual Flight Rules (VFR) are often not on any standard or published routes, the precise path of VFR aircraft is usually unknown or unrecorded by the airport operator. Hence, the aggregate trends of VFR aircraft used for developing noise models are often approximate, as well. Thus, a method of tracking and recording aircraft flight paths is beneficial for use with noise modeling tools to improve the accuracy of the input data, especially for VFR aircraft and aircraft at non-towered airports. Rather than using estimates, approximations, and standard/published procedures (for flight under Instrument Flight Rules (IFR)), the actual tracks of aircraft can be used as input. The increased knowledge and accuracy of aircraft flight paths will enhance the accuracy of noise estimates. This effect is magnified given that most noise metrics are average-based, and an incorrect assumption of the typical flight path can lead to significantly different metrics calculated for a specific position. Further, the knowledge of the actual aircraft used, rather than a predicted aircraft, can ensure that the software uses the best model for that aircraft.

Automatic Dependent Surveillance-Broadcast (ADS-B)

Automatic Dependent Surveillance-Broadcast (ADS-B) is an aircraft surveillance system introduced as part of the FAA's NextGen program, designed to supplement, and improve current aircraft surveillance methods. The implementation of ADS-B has enabled new technologies in areas such as aircraft collision avoidance and operation counts by reducing the costs of locating and monitoring aircraft (Kunzi & Hansman, 2013; McNamara, et al., 2016). Costs are reduced due to the design of the system; equipped aircraft have transponders that transmit position information at least once per second while airborne and at least once per five seconds while on the ground, which are then received by ground stations on the 1,090 or 978 MHz frequency, eliminating the need for RADAR systems to detect and estimate the position of different aircraft (Automatic Dependent Surveillance-Broadcast (ADS-B) Out equipment performance requirements, 2010). While operating ADS-B transponders is not required for all aircraft, they are required for aircraft flying in and around most types of controlled airspace in the United States and are also required in multiple international airspaces (Aircraft Owners and Pilots Association, 2015). Thus, most aircraft, especially those that operate in busier environments, are equipped with ADS-B. As of August 1, 2021, 155,471 out of the approximately 220,000 aircraft (around 70%) in the US are equipped with ADS-B transponders (Bureau of Transportation Statistics, 2020; FAA, n.d.-b). US regulations require the accuracy of ADS-B systems utilized for aerial navigation to be within 0.05 nautical miles (approximately 300 feet) and have a maximum transmission latency of 0.6 seconds if uncompensated, and 2.0 seconds when compensated through extrapolation (Automatic Dependent Surveillance-Broadcast (ADS-B) Out equipment performance requirements, 2010).

ADS-B transponders transmit aircraft identification, surface position, airborne position, airborne velocities, aircraft status, target state and status information, and aircraft operation status, and can be received by an appropriately configured receiver (Sun, 2021). Using aircraft identification, the corresponding aircraft type and engine, relevant to noise models, can be identified using information from the FAA's aircraft registry. Position information transmitted is derived from the Global Navigation Satellite System (GNSS), while altitude data transmitted include both GNSS-derived and barometrically derived altitude. ADS-B transmissions also include a navigation uncertainty category (NUC) with regard to position and velocities, with codes corresponding to different levels of uncertainty bounds. The NUC information can be used to exclude data when not expected to meet the desired specifications.

By using easily and inexpensively collectible ADS-B data, airport operators, especially of non-towered airports, can collect aircraft trajectory data for use with noise estimation models and software such as AEDT. This method potentially offers significantly improved local noise monitoring without the need for expensive data collection tools such as community sound level meters.

To test the concept, a pilot study funded by an Airport Cooperative Research Program Next Step award was conducted at the Purdue University airport. The study was divided into two stages. The first stage, reported in this article, focused on validating and measuring the accuracy of ADS-B data and was intended to answer the research question: how accurate is ADS-B data from one receiver? The second stage, described elsewhere, will report the accuracy of the noise estimates using the collected ADS-B data in conjunction with a simplified noise propagation model in comparison to physically recorded sound levels.

Methodology

Data Collection

To collect ADS-B data for use with the noise estimation method, an ADS-B receiver at the airport being studied was repurposed. The ADS-B receiver used was a Raspberry Pi 3B with a software-defined radio (SDR) connected to a dipole antenna mounted on the roof of the airport's terminal. The receiver utilized the open-source dump1090 software to decode received signals which were then saved using socket30003 software and regularly uploaded to a cloud storage network (Adsbxchange, 2018; Sluis, 2017). The receiver's amplifier gain was adjusted to ensure that ADS-B signals from aircraft within 10 nautical miles of the airport were consistently being received and were neither too weak to be decoded nor too strong that they were filtered out by the software.

To validate the accuracy of the four-dimensional (longitude, latitude, altitude, and time) ADS-B data collected by the receiver, data from 13 G1000NXi units of Purdue University Piper Archer training aircraft were extracted for comparison to serve as the baseline. The G1000NXi Integrated Flight Deck is an avionics package manufactured by Garmin that is comprised of displays, computers, and sensors, replacing traditional analog aircraft instruments and systems.

Beyond providing flight and navigation data for the pilots of the aircraft, the G1000NXi also contains a flight data logging system that records flight and engine data onto an SD card every second. This data contains information including, but not limited to, the local date and time (to the nearest second), UTC offset, latitude (degrees), longitude (degrees), indicated barometric altitude (feet, as adjusted by the altimeter setting) and GPS altitude (feet above mean sea level [MSL]). Each log produced by the system begins as the system boots up and ends when the system is shut off and is periodically uploaded to an online repository via a cellular device installed in the aircraft. Each recorded log usually corresponds to one continuous flight from when the avionics master switch is turned on until it is turned off. This one "flight" may include more than one takeoff and landing such as when "touch and goes" are performed or when landings at multiple airports occur.

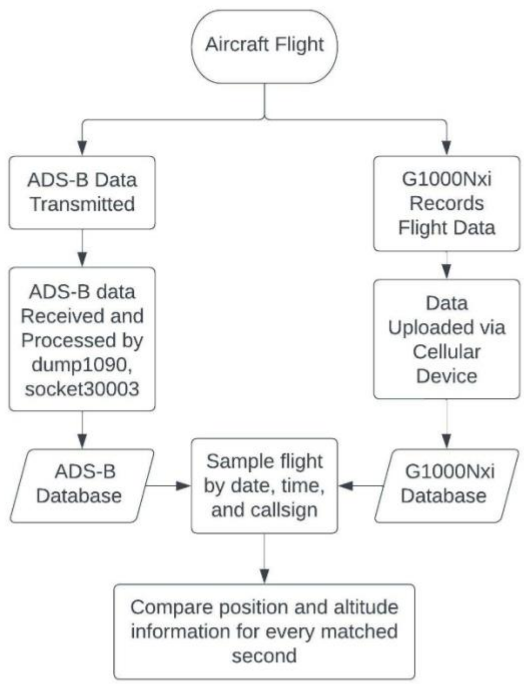
Validation Process

Software was developed and used to calculate the mean error between ADS-B and G1000NXi positions at a given time, measured as the distance or difference in altitude between both reported points, in feet. For every G1000NXi log entry, a corresponding ADS-B record for the date and time was selected, and if more than one ADS-B record for a given second of G1000NXi entry was found, the entry closest to the whole-second (rounded down) was used for the comparison (see Figure 2).

To find the distance between the reported latitude and longitude coordinates, the geodesic distance (based on the WGS-84 ellipsoid model of the earth) was calculated based on an algorithm provided by Karney (2013). While this method assumes the two positions are on the surface of the earth, the altitude of the aircraft was insignificant due to its range (no more than 12,000 feet). This distance based on the latitude and longitude coordinates of the two different records is considered the "lateral" position error.

The vertical (altitude) error and vertical (altitude) absolute error were calculated separately and were provided based on the two different altitude fields recorded by the G1000NXi (indicated barometric altitude and GPS altitude above MSL). The raw error values were found to show the potential positive or negative bias between the ADS-B-reported barometric altitude and the G1000NXi-reported altitude, while the absolute error was also calculated to quantify the error in general. GNSS-derived ADS-B altitude was not used for comparison; only ADS-B-reported barometric altitude was used as that was the only altitude data collected by the utilized software.

Figure 2
Flowchart of ADS-B Data Validation Process



Results

Approximately 3,400 G1000NXi logs from June 3, 2021, to October 31, 2021, were available on the online repository of the Purdue University fleet. A total of 77 random samples were chosen using a randomized index from across the 3,400 total logs. Because 17 of the randomly selected logs had issues preventing their processing (including ground operations not received by the ADS-B receiver, unusable time data due to the limitations of the error-calculating program, or unknown data errors), only 60 of the logs were utilized for comparison.

The 60 sampled logs were from all 13 Piper Archers in the Purdue fleet and included data from June 4, 2021, to October 31, 2021. The sampled logs also encompassed all types of regular movements of university aircraft, including touch and goes, practice maneuvers, and cross-country flights, with an average duration of 78.53 minutes per flight. A total of 246,349 ADS-B and G1000NXi points from the 60 samples used were tabulated and compared to infer the parameters of the lateral and vertical errors.

The sample lateral error had a mean of 57.72 feet with a standard deviation of 44.66 feet, a minimum of 0.07 feet, a median of 47.47 feet, and a maximum of 471.17 feet. With 95% confidence, the mean of lateral errors in the studied population was between 57.55 and 57.90 feet.

The sample absolute vertical error based on the indicated altitude recorded by the G1000NXi had a mean of 112.36 feet with a standard deviation of 69.55 feet, a minimum of 0.00 feet, a median of 121.00 feet, and a maximum of 324.90 feet. With 95% confidence, the mean absolute vertical error based on the indicated altitude recorded by the G1000NXi was between 112.09 and 112.64 feet.

The sample absolute error based on the GPS altitude reported by the G1000NXi had a mean of 142.83 feet, with a standard deviation of 92.92 feet, a minimum of 0.00 feet, a median of 141.10 feet, and a maximum of 462.30 feet. With 95% confidence, the mean raw vertical error based on the GPS altitude reported by the G1000NXi was between 142.46 and 143.20 feet.

The sample raw vertical error based on the indicated altitude parameter had a mean of -81.66 feet with a standard deviation of 103.90 feet, a minimum of -324.90 feet, a median of -109.90 feet, and a maximum of 280.40 feet. With 95% confidence, the mean raw vertical error based on the indicated altitude recorded by the G1000NXi was between -82.07 and -81.25 feet. Based on these data, the G1000NXi reported altitudes above the ADS-B reported altitudes more often than below the ADS-B reported altitudes.

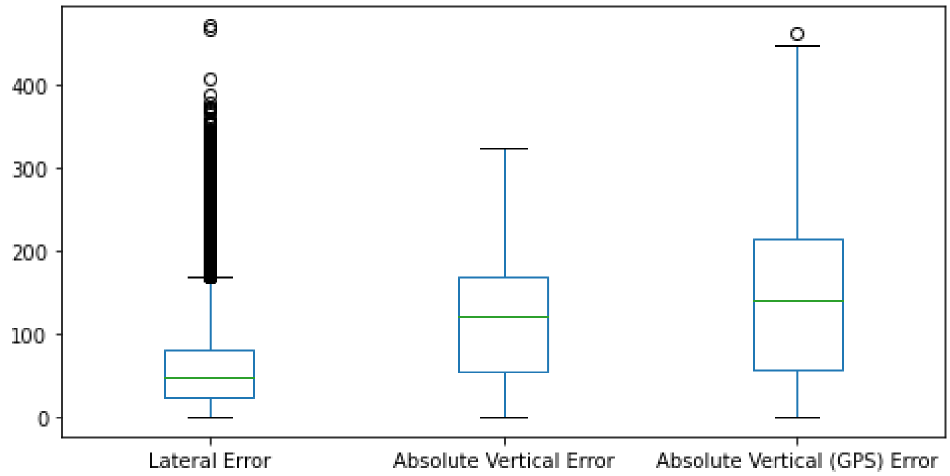
The sample raw vertical error based on the GPS-reported altitude had a mean of -124.23, with a standard deviation of 116.63, a minimum of -462.30 feet, a median of 134.00 feet, and a maximum of 352.20 feet. With 95% confidence, the mean raw vertical error based on the GPS altitude reported by the G1000NXi was between -124.70 and -123.77 feet.

A summary of the sample statistics can be found in Table 1 while boxplots comparing the lateral and absolute vertical errors can be seen in Figure 3.

Table 1
Summary Statistics of Errors

	Mean	Standard Deviation	Minimum	Median	Maximum
Lateral error	57.72	44.66	0.07	47.50	471.17
Absolute vertical error (indicated)	112.36	69.55	0.00	121.00	324.90
Absolute vertical error (GPS)	142.83	92.92	0.00	141.10	462.30
Vertical error (indicated)	-81.66	103.90	-324.90	-109.80	280.40
Vertical error (GPS)	-124.23	116.63	-462.30	-134.40	352.20

Figure 3
Boxplots of Lateral and Absolute Vertical Errors (in feet)



Discussion

Based on the sample data and inferences about the population means, the vertical error between the ADS-B barometric data usually reported lower altitudes than the indicated or GPS altitude recorded by the G1000NXi. Additionally, the error between the ADS-B-reported altitude and the G1000NXi-reported GPS altitude was greater and had higher variability than the error between the ADS-B-reported altitude and the G1000NXi-reported indicated altitude. The lateral error was less than the vertical errors and had lesser variability but with numerous outliers.

Some error is due to a time precision issue as data from the G1000NXi is reported to the nearest second while ADS-B data are reported to the nearest millisecond, and the error calculation program utilized for this project compares data points reported closest to the same second, rounded down. Assuming a groundspeed of 100 knots, an aircraft would travel approximately 169 feet in a given second, allowing an inherent error to exist through this validation procedure even if the data were perfectly accurate. However, errors in altitude, especially when the aircraft is not climbing, or descending, cannot be explained by a precision error in time. Rather, some of the altitude error is due to a precision error in the reported altitude. ADS-B data contains altitude reported in 25-foot increments while the G1000NXi records altitude to the tenth of a foot. Further, the wide spread of error between the recorded ADS-B altitudes and G1000NXi indicated altitudes (altitude adjusted according to the altimeter setting) is due to the ADS-B altitude not being adjusted for non-standard altitude and temperature. This systematic error was previously identified by Mott et al. (2020) who developed a regression model to accurately compensate for the unadjusted ADS-B altitude data. This issue is also probably compounded by calibration and accuracy issues of reporting transponders. Lastly, some error is likely due to the delay between the time the position data is transmitted from the aircraft's transponder to the time it is received and processed by the ADS-B receiver. This issue is most pronounced for aircraft flying further away from the receiver.

Laterally, the mean error of 57.72 feet is about 1.65 times the wingspan of a Piper Archer and it takes approximately 0.34 seconds for an airplane to fly at a groundspeed of 100 knots (assuming it is in the same axis as the error). This calculated error is within the maximum inherent error of the calculation method (one second of flight, approximately 169 feet) and may completely be accounted for by this systematic error. The altitude error could also be further reduced by implementing corrections for atmospheric pressure as described by Mott et al. (2020).

The overall error of the reported ADS-B data based on the G1000NXi data can be described by calculating the resultant "slant" distance from the lateral and vertical errors. Assuming the lateral and vertical errors are independent, the mean slant distance between reported positions can be approximated using the Pythagorean theorem as the square root of the product of the lateral and vertical errors squared, further assuming that any errors resulting from the curvature of the earth are insignificant. Based on this method, the mean slant distance between the ADS-B reported points and the G1000NXi data, using the reported indicated altitude, is 126.32 feet. This result is consistent with the 33-meter (approximately 108 feet) minimum error found in a previous study by Zhang et al. (2011). In the context of noise estimation, this error can lead to an estimation error of up to seven decibels at distances between 200 and 400 feet based on data from the EUROCONTROL NPD database for PA28 aircraft (EUROCONTROL, 2020). At greater distances, due to the logarithmic function of sound levels, this position error would lead to lower noise estimate errors. Given that the lateral error is most likely attributable to a systematic error in the comparison while the altitude error is most likely due to systematic errors from the received ADS-B data, implementing the atmospheric adjustments for the received ADS-B data should be the first step in improving the ADS-B data for noise modeling purposes.

Assumptions and Limitations

A few assumptions and limitations were present in this study. First, given that G1000NXi data were used as the point of comparison for the first stage validation of the ADS-B data, it was assumed that the recorded G1000NXi position and time data were true and accurate. Given that ADS-B and the G1000NXi are GNSS-based systems, the accuracy of the position data will be contingent upon the accuracy and operability of GNSS services during the operations being studied. Second, as no meteorological data was collected as part of this validation process, adjustments to the altitude reported by ADS-B could not be applied nor tested. Additionally, the G1000NXi and ADS-B data came from only one aircraft make, type, and model. The equipment (avionics and transponder) of all the aircraft studied were identical, and as such, errors resulting from the use of different equipment are not captured by the study. The results from this study are representative of Piper Archer aircraft equipped like Purdue University's. Reporting errors (position and latency) of different ADS-B reporting transponders may be different given the higher margin of error accepted by regulations, and as such may warrant further study.

Conclusions

The availability and use of accurate and precise aircraft position data obtained through the operation of cost-effective ADS-B receivers will enable non-towered airports to use noise estimation models and software to estimate airport noise impacts. The historical data provided by ADS-B may also lead to improved noise exposure models when compared to current methods, especially for VFR traffic, by eliminating the need to estimate aircraft flight paths. This article reported calculated ADS-B position errors in comparison with position data from onboard avionics equipment. With an average position error of 57.72 feet laterally, 112.36 feet vertically, and 126.32 feet combined, errors associated with raw ADS-B data should result in noise estimation errors of less than seven decibels at the closest distances (EUROCONTROL, 2020). Position information can be most significantly improved by accounting for atmospheric conditions and adjusting ADS-B reported altitude.

The findings of this study provide additional information on the performance and accuracy of ADS-B. This information can provide an idea of the margin of error associated with ADS-B data use for applications such as those by Huynh et al. (2022) and Gagliardi et al. (2017).

The application of ADS-B in noise modeling can be used in noise impact studies and noise abatement programs at airports and their communities around the US and the world. Further, with the push towards the increased use of air transportation through the integration of Unmanned Aerial Systems, Urban Air Mobility, and Advanced Air Mobility, the same model and concepts from this application can be used to study the noise impacts of such systems. Lastly, the findings related to the accuracy of raw ADS-B data and the need for compensation for atmospheric conditions can be used to guide the development of other uses of ADS-B data.

References

Adsbxchange. (2018). *dump1090-fa* [Computer Software]. https://github.com/adsbxchange/dump1090-fa

Aircraft Owners and Pilots Association. (2015, May 15). Retrieved August 5, 2021, from https://www.aopa.org/go-fly/aircraft-and-ownership/ads-b/where-is-ads-b-out-required

Appendix A to Airport Noise Compatibility Planning, 14 C.F.R. § A36. (2021, August 3).

Appendix A to Noise Standards, 14 C.F.R. § A150. (2021, August 3).

Automatic Dependent Surveillance-Broadcast (ADS-B) Out equipment performance requirements, 14 C.F.R. § 91.227. (2010, June 30).

Bureau of Transportation Statistics. (2020). *Active U.S. air carrier and general aviation fleet by type of aircraft.* Retrieved August 5, 2021, from https://www.bts.gov/content/active-us-air-carrier-and-general-aviation-fleet-type-aircraft

European Organisation for the Safety of Air Navigation. (2020). *The aircraft noise and performance (ANP) database)* (ANP v2.3) [Data set]. https://www.aircraftnoisemodel.org/home

Federal Aviation Administration. (n.d.-a). *Aviation environmental design tool*. Retrieved August 5, 2021, from https://aedt.faa.gov/

Federal Aviation Administration. (n.d.-b). *Current equipage levels*. Equip ADS-B. Retrieved August 5, 2021, from https://www.faa.gov/nextgen/equipadsb/installation/current_equipage_levels

Federal Aviation Administration. (2019, November 4). *Integrated noise model (INM)*. Retrieved August 5, 2021, from https://www.faa.gov/about/office_org/headquarters_offices/apl/research/models/inm_model/

Federal Aviation Administration. (2021a). *Aviation environmental design tool (AEDT) version 3d user manual.* https://aedt.faa.gov/Documents/AEDT3d_UserManual.pdf

Federal Aviation Administration. (2021b). *Aviation environmental design tool (AEDT) version 3d technical manual.* https://aedt.faa.gov/Documents/AEDT3d_TechManual.pdf

Gagliardi, P., Fredianelli, L., Simonetti, D., & Licitra, G. (2017). ADS-B system as a useful tool for testing and redrawing noise management strategies at Pisa Airport. *Acta Acustica united with Acustica*, *103*(4), 543-551.

Huynh, J. L., Mahseredjian, A., & John Hansman, R. (2022). Delayed Deceleration Approach Noise Impact and Modeling Validation. *Journal of Aircraft*, 1-13.

International Civil Aviation Organization. (n.d.). *Environmental technical manual volume 1 procedures for the noise certification of aircraft, 1st ed*. https://www.icao.int/environmental-protection/Documents/Publications/Doc_9501_Volume_I.pdf

Karney, C. F. (2013). Algorithms for geodesics. *Journal of Geodesy*, *87*(1), 43-55.

Kunzi, F., & Hansman, R. J. (2013). *Development of a high-precision ADS-B based conflict alerting system for operations in the airport environment*. https://dspace.mit.edu/bitstream/handle/1721.1/82031/PhDThesis-Kunzi%28ICAT%29.pdf?sequence=1&isAllowed=y

Manci, K. M., Gladwin, D. N., Villella, R., & Cavendish, M.G. (1988, June). Effects of aircraft noise and sonic booms on domestic animals and wildlife: a literature synthesis. U.S. Fish and Wildlife Service, Fort Collins, CO, USA. Retrieved August 5, 2021, from https://www.nonoise.org/library/animals/litsyn.htm#3.1%20MAMMALS

McNamara, M., Mott, J., & Bullock, D. (2016). Leveraging aircraft avionics for fleet and airport management. *Transportation Research Record: Journal of the Transportation Research Board*, *2569*(1), 32–41. https://doi.org/10.3141/2569-04

Mott, J. H., Yang, C., & Bullock, D. M. (2020). Atmospheric pressure calibration to improve accuracy of transponder-based aircraft operations counting technology. *Journal of Aviation Technology and Engineering*, *9*(2), 35.

Pennsylvania State University. (n.d.). *Noise basics*. PSU Noisequest. Retrieved August 5, 2021, from https://www.noisequest.psu.edu/noisebasics.html

Sluis, T. (2017). *dump1090.socket30003* [Computer software]. https://github.com/tedsluis/dump1090.socket30003#dump1090socket30003

Sun, J. (2021). *The 1090 Megahertz Riddle: A Guide to Decoding Mode S and ADS-B Signals (2nd ed.).* TU Delft OPEN Publishing.

Swift, H. (2010, July). *A review of the literature related to potential health effects of aircraft noise.* Partnership for AiR Transportation Noise and Emissions Reduction. https://rosap.ntl.bts.gov/view/dot/28419

Yang, C., & Mott, J. H. (2022). Developing a Cost-Effective Assessment Method for Noise Impacts at Non-Towered Airports: A Case Study at Purdue University. *Transportation Research Record*. https://doi.org/10.1177/03611981221097705

Zhang, J., Wei, L. I. U., & Yanbo, Z. H. U. (2011). Study of ADS-B data evaluation. *Chinese Journal of Aeronautics*, *24*(4), 461-466.

Collegiate Aviation Review International

Volume 40 | Issue 2 Peer Reviewed Article #3

11-1-2022

Exploring Personality and Stress during Communication Delays in Simulated Spaceflight Missions

Shayan Shirshekar
Florida Institute of Technology

Brooke Wheeler
Florida Institute of Technology

Tianhua Li
Florida Institute of Technology

To date, there are no studies exploring how an individual's unique personality profile predicts their response to the stresses and challenges of communication delays on Long Duration Exploration Missions (LDEMs). When exploring astronaut selection for future LDEMs, the Big Five personalities have been identified as a relevant model of personality and one of the preferred models among NASA scientists. This study examined whether personality predicts stress levels when experiencing communication delays during simulated spaceflight missions. A predictive correlational design explored the relationship between personality and stress levels while experiencing a 2-minute one-way communication delay during a simulated Mars mission. Personality included the Big Five personality traits (extraversion, agreeableness, conscientiousness, neuroticism, and openness to experience) and locus of control (LOC). Stress levels were reflected by the difference in stress (DS) scores measured using a stress Visual Analog Scale (VAS). There were significant relationships between conscientiousness and extraversion, both of which were significant predictors of DS scores. LOC was also significantly associated with DS scores. Conscientiousness and extraversion predicted stress when experiencing communication delays. LOC was also identified as a predictor of stress levels. These findings benefit the characterization of crew selection and composition of future spaceflight teams. They also promote a multi-trait, multi-method approach to astronaut selection.

Recommended Citation:

Shirshekar, S., Wheeler, B. & Li, T. (2022). Exploring personality and stress during communication delays in simulated spaceflight missions. *Collegiate Aviation Review International*, 40(2), 33-48. Retrieved from http://ojs.library.okstate.edu/osu/index.php/CARI/article/view/8675/8430

Introduction

Scientists and experts have established that the mission control center (MCC) dynamic of future long-duration exploration missions (LDEMs) will be significantly different from what it is today. Missions to the Moon, Near-Earth orbit, and Mars will present logistical challenges, such as delayed communication between ground control and crew, ranging from a couple of seconds (s) to up to 20 minutes (Palinkas et al., 2017). A human expedition to Mars will involve increasing communication delays between the crew and MCC, rendering normal voice communication impossible within weeks of departing Earth. As crews venture deeper into outer space, the pre-established relationship between the crew and ground/MCC is critical, as they will be the only lifeline crews will have back to Earth. Under these circumstances, effective and clear communication between crewmembers and MCC will be essential for completing mission objectives and maintaining mission safety (Kanas & Manzey, 2008). Without the appropriate support from MCC, communication delays in future missions could negatively impact crewmember performance and behavior (Kanas & Manzey, 2008). Due to the lack of LDEMs to date, there is limited knowledge of how communication delays impact individual crewmembers, and the necessary countermeasures for their mitigation still remain largely unknown (Kintz & Palinkas, 2016).

To date, no studies have explored communication delays from an individual differences perspective, specifically exploring how an individual's personality profile predicts their response to the stresses and challenges of asynchronous communication of varying delays. An initial protocol for an ISS study (Palinkas et al., 2017) investigating the effect of communication delays on well-being and performance planned to include personality traits such as LOC as moderating variables; however, this was ultimately forgone due to concerns expressed by the astronaut office (AO) about the astronaut's willingness to answer specific types of information in a standardized form (Palinkas et al., 2017). Palinkas et al. (2017) would have been the first study to investigate this dynamic. One way to determine which applicants may be unqualified or unsuitable for future deep exploration missions is by measuring different personality constructs.

LOC is defined as an individual's perceived level of control over their situation and experiences that shape their lives (Rotter, 1966). Individuals with an internal LOC believe that events result from their own behavior and actions, whereas individuals with an external LOC believe that events result from an external environmental factor. Although less studied in the spaceflight domain, LOC is a popular construct with important implications in both aviation and military research (e.g., Hunter, 2002; Hunter & Stewart, 2012; You, Ji, & Han, 2013). In a review of predictors and other factors that could contribute to behavioral maladjustment and psychiatric conditions in future spaceflight missions, locus of control was described as one of the components necessary for resilience (ability to sustain or bounce back from different stressors). Specifically, resilient crewmembers of future missions would be those who possess an internal locus of control (Slack et al., 2016). One of the few spaceflight studies exploring LOC was conducted at the Mars-500 experiment, where Russian cosmonauts spent 520 days in group isolation and confinement (Solcova & Vinokhodova, 2015). Among the many conditions simulating LDEM, cosmonauts experienced communication delays varying from 8 to 736 seconds with MCC. When comparing baseline data with follow-up data, LOC was found to become more internal in four of the five cosmonauts, which was thought to reflect personal growth (Solcova & Vinokhodova, 2015). Given cosmonauts' preference for working autonomously, it is expected that individuals with more internal LOCs would be more effective at coping with periods of asynchronous communication in LDEM compared to those with more external LOCs.

Among other existing personality theories and models, one of the preferred models for astronaut selection is Costa and MacCrae's (1992) Five-Factor Model (Landon et al., 2017), also referred to as the Big Five personalities. This theoretical model proposed that five core personality traits serve as the building blocks of personality: extraversion, agreeableness, conscientiousness, neuroticism, and openness to experience. Extraversion (the opposite of introversion) is the level of sociability and outgoingness. Agreeableness is characterized as the degree of interpersonal trust and altruism. Conscientiousness refers to the level of sensibility in decision-making, organization, and self-discipline. Neuroticism is an individual's emotional stability and tendency to experience psychological distress. Finally, openness to experience is proactively seeking to try new things (Costa & McCrae, 1992). In addition, each personality trait represents a range between two extremes. For example, most individuals will score on a continuum somewhere between extraversion and introversion. Furthermore, the research suggested that the five factors are relatively stable and endure throughout adulthood (McCrae & Costa, 2003).

The literature has validated the Big Five personality traits as predictors of performance in a variety of settings, including both organizational employment and isolated, confined, and extreme environments. For example, in a quantitative review summarizing the results of 15 meta-analytic studies investigating the relationship between the Five-Factor Model and job performance, all five personality traits predicted job performance in some way (Barrick et al., 2001). The Big Five personalities have also been researched in aviation, where pilots were found to possess similar personality profiles to astronauts (Fitzgibbons et al., 2004). Most NASA astronauts tend to have some level of piloting experience. This is not to suggest that every member of a future Mars mission will possess extensive piloting experience if any, but that it is an obvious advantage for selection.

In addition to organizational performance, the Big Five personalities have also been explored in isolated, confined, and extreme environments such as Antarctica. For example, in Antarctic analogs, up to 19% of the variance in individual performance was explained by the five personality factors (Palinkas et al., 2000), and neuroticism/emotional stability and agreeableness were found to be the strongest predictors of team performance within the current LDEM literature (Landon et al., 2017). Despite these assertions, it is important to highlight that some studies have yielded contrary results. For example, in a study in which personality data from 259 participants in NASA's final stage of astronaut applicants were collected between 1989 and 1995, results using an abbreviated version of the NEO Five-Factor Inventory (NEO-FFI) indicated that personality traits were not a predictor of applicant acceptance into the astronaut corps (Musson et al., 2004). Plausible explanations for these findings included a lack of heterogeneity among the individuals tested and the possibility of some unidentified aspects of personality that were not assessed, but that may have played a role in the final selection (Musson et al., 2004). These conclusions highlight the multifaceted nature of astronaut selection as no single quality or attribute drives final crew selection, but rather many criteria must be met in order to make the final selection.

Nevertheless, there appears to be a consensus on the ideal personality profile for LDEMs. To best cope with the rigors of a LDEM, it is recommended that astronauts should possess above-average scores on conscientiousness and agreeableness, moderate levels of openness to experience, moderately low to moderately high levels of extraversion, and low levels of neuroticism (Suedfeld & Steel, 2000). According to Landon and colleagues, extremely high or low outliers for any personality factor would suggest unsuitability (Landon et al., 2017). The one exception to the rule would be extremely low levels of neuroticism, which would suggest very high emotional stability (Landon et al., 2017). Although the Big Five personality traits have been validated as one of the preferred models of astronaut selection, prior to this work, the unique role of personality in predicting an individual's stress response when experiencing communication delays in LDEM had not been investigated.

The purpose of this study was to examine the predictive effect of personality on stress levels when experiencing communication delays in a simulated space mission. Subjective stress levels were defined as “Difference in Stress” (DS) scores measured by calculating the difference between a pre- and post-stress Visual Analog Scale (VAS; Lesage et al., 2012). The delay was a 2-min one-way delay representing the early Mars transit phase. Individual differences were measured using the Big Five Inventory Scale-44 (BFI-44) (John et al., 1991; John et al., 2008), resulting in the Big Five personality traits: (a) extraversion, (b) agreeableness, (c) conscientiousness, (d) neuroticism, and (e) openness to experience. Locus of Control (LOC) was included in the analysis as an additional predictor and was defined by scores on Rotter’s Internal/External Scale (Rotter, 1966). We hypothesized that the Big Five personality traits and LOC were significant predictors of stress during a 2-minute one-way communication delay in a simulated spaceflight mission.

Method

Population and Sample

The target population for this study was adult males and females in NASA’s astronaut training program who may have been eligible for a future LDEM. The ideal accessible population consisted of members of NASA’s astronaut candidate class; however, International Traffic in Arms Regulations and other restrictions made this unobtainable. For the purpose of this study, the accessible population was all healthy individuals affiliated with a private university in Florida as either students, faculty members, or staff, who possessed a bachelor’s degree or were at minimum junior undergraduate level status. This accessible population was selected as a close approximation of individuals, similar to the ideal population because the members of NASA’s astronaut training program must also be in good health and possess at least an undergraduate degree.

The sampling strategy for this study was convenience sampling. This study also may have inadvertently experienced snowball recruitment as participants who completed the study could have actively recruited colleagues, friends, and roommates. A demographic survey used for recruitment confirmed that the participants met the selection criteria, including the minimum educational requirements and the standards for self-reported astronaut-specific statements related to vision and healthy blood pressure. Finally, though they may not all presently represent the NASA astronaut candidate class, those meeting the selection criteria would meet the minimum requirements to eventually apply for NASA's astronaut training program. This was important in order to provide a more representative sample of the target population.

The demographic survey was made available online via *Qualtrics* and shared over email through the institution's online forum. Recruitment fliers were also put up strategically around the campus with contact information for those interested in participating in the study. The minimum sample size from the a priori power analysis for a sequential regression, with a medium effect size of 0.15 and six predictors, yielded 98 subjects for the study (Faul et al., 2009). A total of 198 individuals submitted responses to the online demographic survey for study participation eligibility. Of those 198 responses, 118 participants met the inclusion criteria, responded to a follow-up email for scheduling, and successfully completed the study. Eighteen responses were excluded for incomplete data, leaving a total of 100 participants with complete data on the dependent variable, the six predictor variables, and all the extraneous variables for further analysis. Participant age ranged from 19 to 51, with a mean of 23.32 and a standard deviation of 5.81. Also, 42 participants were male, whereas 58 were female. Finally, 31 of the participants reported having some level of piloting experience.

Apparatus

Stress Visual Analog Scale (VAS)

Stress VASs are commonly used in the medical field by occupational physicians to assess stress among workers (Dutheil et al., 2012; Dutheil et al., 2013; Lesage et al., 2012) and have been suggested as a tool for assessing perceived stress in both clinical and research settings (Lesage et al., 2012). This computerized assessment consisted of a single item (i.e., "What is your current stress level?") with a horizontal line divided into equally sized partitions on an 11-point scale (0 = No Stress to 10 = Agonizing) and a sliding locator. Participants were asked to mark the point that best represents their perception of their current stress state. This assessment was appropriate because it provided a quick, situational measure of perceived stress. VASs afford rapid administration and high completion rates, providing a useful advantage over more standard multi-item inventories that require more time and effort from the participant (Rossi & Pourtois, 2012). Furthermore, the use of a VAS over a more standard, multi-item inventory might have helped prevent any disruption to the flow of the experiment (MacLeod et al., 2012; Poma et al., 2005). The stress VAS was presented alongside other visual analog scales assessing dimensions including fatigue, level of difficulty, and communication quality. These supplemental scales were included to reduce participant awareness about the variables of interest. Experimental studies have found the stress VAS to possess good sensitivity for stress events, and other work has shown the stress VAS to possess very satisfactory psychometric properties (Lesage et al., 2012).

BFI-44

The Big Five Inventory Scale-44 (John et al., 1991; John et al., 2008) is a 44-item measure with five scales: Extraversion (8 items), Agreeableness (9 items), Conscientiousness (9 items), Neuroticism (8 items), and Openness to experience (10 items). Participants are provided the phrase: "I am someone who…", followed by an item statement that participants rate in terms of agreement using a 5-point Likert scale ranging from 1 (*Disagree Strongly*) to 5 (*Agree Strongly*). The scale was developed as a time-efficient alternate measure of the Five-Factor Model that can be completed in approximately 10 min. The BFI-44 has been shown to possess a clear five-factor structure, reliability, convergent validity with other Big Five scales (such as the NEO-PI-R and NEO-FFI), and strong self-peer agreement (Benet-Martinez & John, 1998; John et al., 2008; Soto et al., 2008). The alpha reliability coefficients have been previously reported as .86 for extraversion, .79 for agreeableness, .82 for conscientiousness, .87 for neuroticism, and .83 for openness to experience, yielding an average of .83 (John et al., 2008). Based on the above findings, the BFI-44 was selected for measuring the Big Five personalities in this study for its robust psychometric properties and efficiency.

Rotter's Internal/External Scale

Rotter's Locus of Control Scale has been found to possess acceptable reliability and validity (Goodman & Waters, 1987; Rotter, 1966). This scale consisted of 29 forced-choice items where participants selected one of two options. Following Rotter's guidelines, scores were obtained by adding one point for specific items on the scale and then taking the sum of those scores. Among the 29 questions, items 1, 8, 14, 19, 24, and 27 were filler questions and excluded from data analysis, yielding scores ranging from 0 to 23. High scores reflected an external LOC, and low scores reflected an internal LOC.

Demographic Survey

A demographic survey was used to screen participants for eligibility with respect to educational status and to collect specific demographic information: participants' age, gender, and piloting experience. Participants were also asked to self-report their level of agreement with statements related to vision and blood pressure. The demographic survey was made available online via *Qualtrics*.

Simulation

The simulation was executed using the Re-entry Space Simulator by Wilhelmsen Studios (Wilhelmsen, 2018). This space flight simulation game was made specifically for personal computers and provided a realistic and interactive experience from the viewpoint of an astronaut. It is equipped with access to full historical missions, such as Project Mercury and the Apollo, and thanks to a custom mission editor, allowed the creation of new missions designed to challenge the user. All virtual cockpits were fully interactive, allowing the user to manually control almost every single component of the cockpit, from gauging electrical systems to environmental control. For the purpose of this study, many of the elements from the Mercury capsule mission module were customized and developed to represent a Mars mission. The Mercury project contained some of the most basic controls and configurations ideal for this study. This was redesigned to make the participant feel fully immersed in a Mars mission, including modified backdrop graphics showing the red planet while completing tasks, system checks, and emergency protocols all characteristic of a real Mars mission. Prior to developing the customized mission, Re-entry was carefully inspected by multiple subject matter experts (SMEs) to ensure it possessed the necessary capabilities for carrying out simulated missions with communication delays as well as providing an adequate level of fidelity with respect to realism. Both concerns were met prior to developing the customized missions.

Study Design

This study used a correlational design to explore the predictive effect of the Big Five personalities on stress levels. There were two mission emergencies: primary life-support systems (PLSS) pressure regulator failure and carbon dioxide (CO_2) scrubber failure. These emergencies were evaluated by SMEs and deemed equivalent, and half of the participants were assigned to each. The basic structure of the 45-min mission was similar for each participant. Using re-entry, the mission incurred a 2-min one-way delay that would be expected during the initial stages of the transit phase (approximately 30 days into the mission). Participants were first briefed about the study and protocols. The briefing was read from a script and helped participants familiarize themselves with the setting, audio and communication equipment (i.e., walkie-talkie), and input controls for operating Re-entry (i.e., mouse and keyboard). Participants were also provided some tips on how to communicate efficiently with MCC during the simulation. Prior to starting, participants were required to complete a stress VAS to record baseline subjective stress scores. After completing the stress VAS, participants began the simulation, finding themselves inside the cockpit of a spacecraft, looking towards the darkness of space.

The simulation started with a 5–10-min tutorial involving artificial intelligence (AI), introducing and welcoming the participant to the mission. The AI then explained the controls and introduced the participant to the cockpit he or she would be operating in. Some of the controls included how to click buttons and switches, pull levers, and turn knobs. The AI also taught the participant how to monitor different panels, including pressure, carbon dioxide, temperature, oxygen, and battery levels. In addition to familiarizing participants with the controls, buttons, and the location of all the panels in the cockpit, the tutorial was also designed to reduce the chance of high-stress levels due to unfamiliarity when participants worked through the simulation.

After the tutorial was completed, the AI notified the participant that the 45-min mission would begin by clicking the prompt on the screen. The mission was divided into two segments, with the first 15 min dedicated to basic tasks, including monitoring systems, attitude control, and photography, and the latter 30 min for responding to one of two potential emergency scenarios programmed to occur at the 15-minute mark of the mission. The emergency scenarios were selected based on the results of Stuster and colleagues' Mars task analysis report (Stuster et al., 2019). They rated "*Respond to technical emergencies, following procedures and with equipment* *provided, during Cruise to Mars"* (p .35) as the most important summary task statement. This summary statement was composed of 15 tasks. From those, three separate emergency scenarios were created: a) primary life-support systems (PLSS) pressure regulator failure, b) carbon dioxide (CO_2) scrubber failure, and c) lighting system power outage. Because this was part of a larger research project (Shirshekar, 2021) that investigated stress levels under delay versus no delay and thus required all participants completing two missions to compare the delay mission with a control, the selected emergency scenarios could not be functionally identical. However, they did need to be objectively similar, such that the number of steps, training and time required to complete each protocol were approximately equivalent. The scenarios also needed to be subjectively comparable by possessing similar levels of perceived risk, difficulty (complexity), and importance to the mission. This was essential to reduce any potential internal validity threats wherein differences in stress levels between the delay and control mission may have been due to slight discrepancies between the two emergency scenarios. Furthermore, by having multiple emergency scenarios, we ensured that any learning effects were minimized. The PLSS pressure regulator failure and CO_2 scrubber failure were rated as the most comparable of the three by SMEs and were selected and pilot-tested prior to this study. Finally, to satisfy the requirements of the larger research project (Shirshekar, 2021), all participants were counterbalanced based on scenario and delay. Thus, half the participants completed the delay scenario under the PLSS pressure regulator failure, while half the participants completed the delay scenario under the CO2 scrubber failure. This paper presents the findings for the delayed mission only.

The first 15 minutes were marked by several transmissions that were prerecorded and delivered by the researcher as MCC using a laptop and a speaker. The transmissions played through the speaker were relayed to the participant via the walkie-talkie to ensure consistency in the sound of transmissions from MCC. Some of the prerecorded transmissions were received by the participant at fixed times, while others were received based on their progress. The remainder of the transmissions were sent live from MCC and not prerecorded. Some of these transmissions were as short as a single word (i.e., roger). Several MCC generic responses to general questions from the participants in the pilot study were compiled and used in the study to minimize the variability of MCC transmissions. The transmissions were also contingent on what the participant was requesting or stating to MCC. For example, one participant might have required more assistance and thus more guidance from MCC, while others might have been more autonomous and less communicative. Thus, the number of transmissions, including both prerecorded and live from MCC and those sent by the participant, varied. In the first transmission, the participant was made aware of the mission pad on the table in front of them containing various checklists that needed to be completed.

Following the 45-min mission duration, the simulation was terminated, and participant post-mission stress levels were immediately collected via the stress VAS. Participants were then asked to complete the BFI-44 (John et al., 1991; John et al., 2008) and the Internal/External Scale (Rotter, 1966). To prevent participants from completing the mission with time to spare and to avoid having them simply wait idly for the 45-min clock to run out, the emergency scenarios were specially designed and tested to require a minimum of 30 min (beginning at the 15-min mark) to complete. As per SME input, this maintained the simulation time as a constant and avoided variable mission durations.

Participants were monitored via video feed by a researcher functioning as MCC in a nearby room. The simulation room was far enough from the researcher to ensure the participant could not hear the researcher, unless it was through their walkie-talkie. The simulation room was designed and configured based on some of the requirements for future LDEMs. One of the main design considerations was the size of the workspace. The researcher followed the guidelines and recommendations for the minimum acceptable Net habitable volume (NHV) for future LDEMs. NHV is defined as the minimum volume of a habitat that is necessary for mission success during LDEM missions with prolonged periods of isolation and confinement in a harsh/extreme environment (Whitmire et al., 2014). According to a consensus on minimum acceptable NHV, a minimum acceptable NHV of 25 m^3 (883 $ft.^3$) is recommended per person for future exploration missions with a maximum duration of 912 days. Furthermore, a workspace of 8.12 m^2 is recommended to allow up to four crewmembers to work simultaneously (Whitmire et al., 2014). The allowable workspace in the simulation room was designed based on these recommendations. Another design consideration was isolation, which includes physical isolation and acoustic isolation. Although it is not possible to re-create the physical isolation of future LDEMs, the participant and researcher were placed in separate rooms, and the simulation room was sound-proofed from outside noise distractions using a surround sound system that played continuous sounds of celestial white noise throughout the duration of the simulation. On the International Space Station, air circulation fans and other equipment produce a constant level of background white noise, making this a realistic soundscape. Finally, to re-create the lack of sensory stimulation and monotonous conditions, the windows were covered with two large projector screens. Prior to making these configurations, SMEs were consulted to ensure the proper level of fidelity was achieved. Of course, certain conditions such as microgravity and a true sense of isolation were not possible, thus threatening the ecological validity of the study. Within the simulation room, a chair and table were provided, along with the mission pad containing the checklists necessary for completing the mission.

Participants were assigned a number to ensure that the data remained anonymous following the completion of the study. Completion of the study was incentivized with a choice of space-related merchandise, such as socks, t-shirts, and other paraphernalia. The study was approved by the Institutional Review Board (IRB) of a private university in Florida, where the data were collected to ensure that attention was given to human subject research issues.

Results

The stress VAS was administered twice during the study, specifically at the beginning and end of the mission. The mean DS (post-test subtracted from the pretest) score was M = 1.73, with scores ranging from -3–7. The personality measures were collected after the mission was completed. All five of the Big Five personality traits were normally distributed, and the mean for LOC was M = 11.24, with scores ranging from 3–20. The descriptive statistics for the six continuous independent variables and the dependent variable (DS scores) are summarized in Table 1**.**

Table 1
Descriptive Statistics

	N	*M*	*SD*	Minimum	Maximum	Range
DS	100	1.73	2.04	-3	7	10
Extraversion	100	3.33	0.79	1.50	5.00	3.50
Agreeableness	100	3.92	0.59	2.33	5.00	2.67
Conscientiousness	100	3.74	0.63	2.00	5.00	3.00
Neuroticism	100	2.97	0.72	1.25	4.50	3.25
Openness	100	3.97	0.47	2.20	4.90	2.70
LOC	100	11.24	4.16	3	20	17

Note. These are descriptive statistics for the six predictor variables.
M = mean. SD = standard deviation. LOC = locus of control.

Inferential statistics consisted of a sequential multiple regression for DS scores. The predictor variables included extraversion, agreeableness, conscientiousness, neuroticism, openness to experience, and LOC. Multicollinearity was examined first. The variance inflation factors among the predictors were all less than 2. This demonstrated no multicollinearity. Regression assumptions were then examined to avoid biased regression coefficients and standard error estimates. The correction specification of the form of the relationship, reliable measurement of the independent variables, the constant variance of the residuals, independence of residuals, and normality of residuals assumptions were met. However, agreeableness, neuroticism, and openness to experience were removed from the inferential analysis to meet the correct specification of the independent variables.

The objective of this study was to examine the relationship between individual differences in personality and stress levels following a communication delay simulation. A sequential multiple regression was run, with extraversion and conscientiousness in the first model, followed by LOC entering next. The multiple regressions demonstrated a significant regression of extraversion and conscientiousness on DS scores, $F(2, 97) = 3.73$, $p = .03$, $R^2 = .07$, $RMSE = .20$. This indicated that extraversion and conscientiousness predicted 7% of the variance in DS scores. Examining the individual factors in this model, significant relationships were found with extraversion, $t(97) = -2.12$, $p = .04$, $sr^2 = .04$ with $B = -.55$, $\beta = -.22$, $SE = .26$; and conscientiousness, $t(97) = 2.24$, $p = .03$, $sr^2 = .05$ with $B = .74$, $\beta = .23$, $SE = .33$. Interpreting the corresponding regression coefficients, $B_E = -0.55$ and $B_C = 0.74$, these results indicate that by holding all other variables constant, a participant's DS score is predicted to drop approximately 0.55 points on average for every one-point increase in extraversion score in the presence of conscientiousness, and a participant's DS score is predicted to increase by 0.74 points on average for every one-point increase in conscientiousness score in the presence of extraversion.

When LOC entered into the model, the regression was significant, $F(3, 96) = 5.74$, $p < .01$, $R^2 = .15$, $RMSE = .19$. This showed that extraversion, conscientiousness, and LOC predicted 15% of the variance in DS scores. The increase caused by LOC entering the model was also significant, $\Delta F(1, 96) = 9.13$, $p < .01$, $\Delta R^2 = .08$. This indicated that LOC predicted an additional 8% of the variance in DS scores. The parameters are summarized in Table 2. This demonstrated that by holding all other variables constant, a participant's DS score is predicted to drop approximately 0.39 points on average for every one-point increase in extraversion score in the presence of conscientiousness and LOC, and a participant's DS score is predicted to increase by 0.96 points on average for every one-point increase in conscientiousness score in the presence of extraversion and LOC. Also, for every 1-point increase in LOC, the DS score is predicted to increase by .15 points on average in the presence of extraversion and conscientiousness.

Table 2
Sequential Multiple Regression Model Summary.

	B	***SE***	***CI***	**β**	***t***	***p***	***pr2***	***sr2***
(Constant)	-2.21	1.61	[-5.40, 0.98]		-1.37	0.17		
E	-0.39	0.26	[-0.90, 0.12]	-0.15	-1.53	0.13	0.02	0.02
C	0.96	0.33	[0.31, 1.60]	0.29	2.94	0.00	0.08	0.08
LOC	0.15	0.05	[0.05, 0.25]	0.30	3.02	0.00	0.09	0.08

Note. These are inferential statistics for the final regression model with extroversion (E), conscientiousness (C), and locus of control (LOC). *B* = coefficient. *SE* = standard error. *CI* = confidence interval. β represents the standardized regression coefficient of each variable. sr^2 and pr^2 were partial and semi-partial correlation coefficients, which indicated the portion of the variance explained solely by the variable in the presence of other variables and the one by removing the effect of other variables.

Discussion

The findings indicated that two of the Big Five personality traits, extraversion and conscientiousness, were significant predictors of DS scores when incurring a 2-min one-way delay with MCC. Together, extraversion and conscientiousness predicted about 7% of the variance in DS scores. LOC was also a significant predictor of DS scores, predicting another 8% of the variance. These results may collectively imply that conscientiousness, extraversion, and LOC could be important personality traits for future LDEM crew selection. For example, increased levels of conscientiousness among crewmembers may lead to increased awareness of communication delays during LDEMs, thus resulting in higher stress levels. Similarly, higher LOC scores are indicative of a more external disposition. Thus, crewmembers with more external LOC scores could be more reliant on outside help (i.e., MCC), which could be challenging with the implemented delay. Although increased levels of extraversion correlated with lower stress levels, it could be theorized that crewmembers with excessively high levels of extraversion, who would likely be in more need of conversation and social interaction, could be prone to feeling increased stress and frustration from the limited communication with MCC. Thus, some degree of introversion may be desirable (Landon et al., 2017). On the other hand, given that only two of the Big Five personality traits were significant predictors, these findings may also collectively imply that the Five-Factor Model may not be a major predictor of individual well-being when incurring a communication delay with MCC, and that other personality models or traits, such as LOC, may merit further investigation. This would support the current multifaceted practice of NASA astronaut selection, in which no single factor drives selection. This implication is also timely with the rise of private companies and space tourism, where everyday individuals possessing unique personality profiles may be selected for future space missions.

One recommendation for practice is for NASA and private spaceflight companies to place further emphasis on Big Five traits, such as conscientiousness and extraversion as well as other personality traits such as LOC. However, because of the lack of astronaut performance data to date, it is difficult to draw firm conclusions for astronaut selection recommendations. It is recommended that future researchers replicate the study but increase the length of both the communication delay and the simulation. Previous literature exploring communication delays in behavior and performance suggest at least 1 hour is necessary to ensure enough time to capture behavioral assessments and complete ratings (Palinkas et al., 2017). Another recommendation is that researchers study individual well-being and performance in tandem, as much of the literature exploring the predictive effect of the Big Five in isolated, confined, and extreme conditions are predominantly performance-based. It is possible individual well-being and performance may be associated with one another or that performance could mediate the relationship between delay and individual well-being. This study was limited in its capacity to implement multiple stress assessment tools due to the COVID-19 Pandemic. It is recommended that future studies implement multiple stress measures to capture a more comprehensive evaluation of stress. This could include both subjective and physiological assessments. Because stress is a multidimensional construct, assessing multiple physiological markers (blood pressure, heart rate, galvanic skin response, and salivary markers) may be more useful for documenting effects. Furthermore, previous studies investigating VAS data have found them to lie somewhere in between ordinal and interval scales (McCormack et al., 1988; Philip, 1990; Price et al., 1994). However, with scores generally ranging from 0 to 100 (or 0 to 10), they are said to have equality between intervals and can thus be subjected to parametric statistics (Kersten et al., 2014). Following these assertions, it was decided that the DS scores would be considered interval scale for this study, but it must be cautioned that the literature remains somewhat ambiguous as to where VAS data truly lie.

This was the first study to explore the relationship between personality and stress during communication delays; to build on this work and prepare for potential missions, future communication delay studies should investigate stress measures and personality from a team perspective. Finally, it is recommended that future researchers incorporate more predictor variables to reflect the holistic multi-trait, multi-method approach, as no single factor drives astronaut selection.

References

Barrick, M. R., Mount, M. K., & Judge, T. A. (2001). Personality and performance at the beginning of the new millennium: What do we know and where do we go next? *International Journal of Selection and Assessment*, *9*(1-2), 9-30. https://doi.org/10.1111/1468-2389.00160

Benet-Martinez, V., & John, O. P. (1998). Los Cinco Grandes across cultures and ethnic groups: Multitrait multi-method analyses of the Big Five in Spanish and English. *Journal of Personality and Social Psychology*, *75*(3), 729-750. https://doi.org/10.1037/0022-3514.75.3.729

Costa, P. T., & McCrae, R. R. (1992) *NEO PI-R professional manual*. Psychological Assessment Resources Inc.

Dutheil, F., Boudet, G., Perrier, C., Lac, G., Ouchchane, L., Chamoux, A., Duclos, M., & Schmidt, J. (2012). JOBSTRESS study: Comparison of heart rate variability in emergency physicians working a 24-hour shift or a 14-hour night shift — A randomized trial. *International Journal of Cardiology*, *158*(2), 322-325. https://doi.org/10.1016/j.ijcard.2012.04.141

Dutheil, F., Trousselard, M., Perrier, C., Lac, G., Chamoux, A., Duclos, M., Naughton, G., Mnatzaganian, G., & Schmidt, J. (2013). Urinary interleukin-8 is a biomarker of stress in emergency physicians, especially with advancing age--the JOBSTRESS* randomized trial. *PloS one*, *8*(8), 1-7. https://doi.org/10.1371/journal.pone.0071658

Dunbar, B. (2021, January 29). *Astronaut Requirements*. NASA. https://www.nasa.gov/audience/forstudents/postsecondary/features/F_Astronaut_Requirements.html

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, *41*(4), 1149. https://doi.org/10.3758/BRM.41.4.1149

Fitzgibbons, A. A., Davis, D., & Schutte, P. C. (2004). *Pilot personality profile using the NEO-PI-R*. NASA/TM-204–213237.

Goodman, S. H., & Waters, L. K. (1987). Convergent validity of five locus of control scales. *Educational and Psychological Measurement*, *47*(3), 743-747. https://doi.org/10.1177/0013164487047003026

Hunter, D. R. (2002). Development of an aviation safety locus of control scale. *Safety*, *7*, 160.

Hunter, D., & Stewart, J. (2012). Safety locus of control and accident involvement among army aviators. *International Journal of Aviation Psychology*, *22*(2), 144-163. https://doi.org/10.1080/10508414.2012.663244

John, O. P., Naumann, L. P., & Soto, C. J. (2008). Paradigm shift to the integrative Big Five trait taxonomy: History, measurement, and conceptual issues. In O. P. John, R. W. Robins, & L. A. Pervin (Eds.), *Handbook of personality: Theory and research (2nd ed.)* (pp. 114-158). The Guilford Press.

John, O. P., Donahue, E. M., & Kentle, R. L. (1991). The Big Five Inventory--Versions 4a and 54. Berkeley, University of California, Berkeley, Institute of Personality and Social Research.

Kanas, N., & Manzey, D. (2008). *Space psychology and psychiatry* (2nd ed.). Springer.

Kersten, P., White, P. J., & Tennant, A. (2014). Is the pain visual analogue scale linear and responsive to change? An exploration using rasch analysis. *PLoS ONE*, *9*(6), 1-10. https://doi.org/10.1371/journal.pone.0099485

Kintz, N. M., & Palinkas, L. A. (2016). Communication delays impact behavior and performance aboard the international space station. *Aerospace Medicine and Human Performance*, *87*(11), 940-946. https://doi.org/10.3357/amhp.4626.2016

Landon, L. B., Rokholt, C., Slack, K. J., & Pecena, Y. (2017). Selecting astronauts for long-duration exploration missions: Considerations for team performance and functioning. *REACH - Reviews in Human Space Exploration*, *5*, 33-56. https://doi.org/10.1016/j.reach.2017.03.002

Lesage, F., Berjot, S., & Deschamps, F. (2012). Clinical stress assessment using a visual analogue scale. *Occupational Medicine*, *62*(8), 600-605. https://doi.org/10.1093/occmed/kqs140

MacLeod, C., Rutherford, E., Campbell, L., Ebsworthy, G., & Holker, L. (2002). Selective attention and emotional vulnerability: Assessing the causal basis of their association through the experimental manipulation of attentional bias. *Journal of Abnormal Psychology*, *111*(1), 107-123. https://doi.org/10.1037/0021-843X.111.1.107

McCormack, H. M., Horne D. J., Sheather, S. (1988). Clinical applications of visual analogue scales: a critical review. *Psychological medicine*, *18*(4), 1007-1019. https://doi.org/10.1017/s0033291700009934

McCrae, R. R., & Costa, P. T. (2003). *Personality in adulthood: A five-factor theory perspective* (2nd ed.). Guilford Press. https://doi.org/10.4324/9780203428412

Musson, D. M., Sandal, G. M., & Helmreich, R. L. (2004). Personality characteristics and trait clusters in final stage astronaut selection. *Aviation, Space, and Environmental Medicine*, *75*(4), 342-349.

Palinkas, L. A., Gunderson, E. K., Holland, A. W., Miller, C., & Johnson, J. C. (2000). Predictors of behavior and performance in extreme environments: the Antarctic space analogue program. *Aviation, Space, and Environmental Medicine*, *71*(6), 619-625.

Palinkas, L.A., Kintz, N., Vessey, W.B., Chou, C.P., & Leveton, L.B. (2017) *Assessing the impact of communication delay on behavioral health and performance: An examination of autonomous operations utilizing the international space station*. National Aeronautics and Space Administration. NASA/TM-2017-219285

Philip, B. K. (1990). Parametric statistics for evaluation of the visual analog scale. *Anesthesia and analgesia*, *71*(6), 710. https://doi.org/10.1213/00000539-199012000-00027

Poma, S. Z., Milleri, S., Squassante, L., Nucci, G., Bani, M., Merlo-Pich, E., & Perini, G. I. (2005). Characterization of a 7% carbon dioxide (CO_2) inhalation paradigm to evoke anxiety symptoms in healthy subjects. *Journal of Psychopharmacology*, *19*(5), 494-503. http://dx.doi.org/10.1177/0269881105056533

Price, D. D., Bush, F. M., Long S., & Harkins, S. W. (1994). A comparison of pain measurement characteristics of mechanical visual analogue and simple numerical rating scales. *Pain*, *56*(2), 217-26. https://doi.org/10.1016/0304-3959(94)90097-3

Rossi, V., & Pourtois, G. (2012). Transient state-dependent fluctuations in anxiety measured using STAI, POMS, PANAS or VAS*: A comparative review. Anxiety, Stress & Coping*, *25*(6), 603-645. https://doi.org/10.1080/10615806.2011.582948

Rotter, J. B. (1966). Generalized expectancies for internal versus external control of reinforcement. *Psychological Monographs*, *80*(1), 1-28. https://doi.org/10.1037/h0092976

Shirshekar, S. (2021). *Exploring the effects of communication delays and personality on stress during simulated space missions* (Doctoral dissertation). https://repository.lib.fit.edu/handle/11141/3368

Slack, K. J., Williams, T. J., Schneiderman, J. S., Whitmire, A. M., Picano, J. J., Leveton, L.B., Schmidt, L. L., & Shea, C. (2016). *Risk of adverse cognitive or behavioral conditions and psychiatric disorders: evidence report.* NASA/Lyndon B. Johnson Space Center. http://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/20160004365.pdf

Solcova, I., & Vinokhodova, A. G. (2015). Locus of control, stress resistance, and personal growth of participants in the Mars-500 experiment. *Human Physiology*, *41*(7), 761-766. https://doi.org/10.1134/S0362119715070221

Soto, C. J., John, O. P., Gosling, S. D., & Potter, J. (2008). The developmental psychometrics of big five self-reports: Acquiescence, factor structure, coherence, and differentiation from ages 10 to 20. *Journal of Personality and Social Psychology*, *94*(4), 718-737. https://doi.org/10.1037/0022-3514.94.4.718

Stuster, J., Adolf, J., Byrne, V., Greene, M. (2019) *Generalizable Skills and Knowledge for Exploration Missions*. NASA Contractor Report-2018-220445

Suedfeld, P., & Steel, G. D. (2000). The environmental psychology of capsule habitats. *Annual Review of Psychology*, *51*(1), 227-253. https://doi.org/10.1146/annurev.psych.51.1.227

Whitmire, A., Leveton, L., Broughton, H., Basner, M., & Kearney, A. (2014). *Minimum acceptable net habitable volume for long-duration exploration missions subject matter expert consensus session report*. NASA/Langley Research Center. NASA TM-2015-218564.

Wilhelmsen, P. (2018). Reentry - An Orbital Simulator [Steam]. Wilhelmsen Studios.

You, X., Ji, M., & Han, H. (2013). The effects of risk perception and flight experience on airline pilots' locus of control with regard to safety operation behaviors. *Accident; analysis and prevention*, *57*, 131-139. https://doi.org/10.1016/j.aap.2013.03.036

Collegiate Aviation Review International

Volume 40 | Issue 2 Peer Reviewed Article #4

11-1-2022

Towards Safer Flight Operations: The Relationship Between L2 Motivation and L2 Achievement

Gökhan Demirdöken
National Defense University

Derin Atay
Bahcesehir University

The purpose of this paper was to identify the relationship between L2 motivation and L2 achievement, a pivotal topic in the aviation industry since it requires more and more proficient English-speaking pilots. Therefore, we aimed to find out the L2 motivational variables that affected L2 achievement in Aviation English courses. The sampling included 111 aviation students. L2 Motivational Self System Questionnaire and Achievement Motivation Measure were instrumented, and semi-structured interviews were conducted as part of our mixed-methods sequential explanatory research design. Our findings revealed a moderately positive correlation between ideal L2 self and L2 achievement; on the flip side, there was a weak negative correlation between ought-to L2 self and L2 achievement. What's more, the L2 achievement of aviation students was found to be predicted by the ideal L2 self and ought-to L2 self; however, achievement motivation failed to account for the L2 achievement of aviation students. The pedagogical implications were discussed in the relevance of L2 motivational variables to aviation students' L2 achievement for an improved learning experience.

Recommended Citation:
Demirdöken, G. & Atay, D. (2022) Towards safer flight operations: The relationship between l2 motivation and l2 achievement. *Collegiate Aviation Review International*, 40(2), 49-66. Retrieved from http://ojs.library.okstate.edu/osu/index.php/CARI/article/view/8710/8431

Study Background

The impetus for this research is the need for a better understanding of L2 motivational variables affecting aviation students' L2 achievement since the aviation industry relies on skilled pilots both in terms of pilotage and English language proficiency. Having already been reported as a contributing factor to a number of aviation accidents, including the Tenerife Disaster, the English language proficiency of pilots is an expanding area of interest within the field of English for Specific Purposes (ESP) which focuses on the purposeful use of the language (Hyland, 2022). There is no doubt that the aviation industry is expanding at a rapid pace, and Aviation English has become the lingua franca for aviation professionals. Besides, it is more and more required by professionals in a particular industry to do more than just polish their styles or control their linguistic errors (Alfehaid & Alkhatib, 2020), and the aviation industry is just one of them. Within this framework, it is of utmost importance that aviation students become proficient speakers of this "codified, abbreviated, jargon-filled" language (Trippe & Baese-Berk, 2019), housing a wide range of language use situations (Mede et al., 2018). Therefore, if the variables predicting L2 achievement in Aviation English can be identified, an improved and effective language learning experience can be offered to this particular group of ESP students. In this regard, we aimed to explore the factors that may predict aviation students' L2 achievement.

Aviation English in Turkish Setting

The growing interest in aviation has paved the way for a significant increase in the number of Aviation English courses offered in pilotage schools and at tertiary-level institutions in Turkey. While there were only two institutions accredited by the Turkish Directorate General of Aviation (DGCA) until 2010, the number of these institutions has increased by 2,350% over the last decade. Today, Aviation English courses are offered by 15 universities and 34 private institutions in Turkey as well as the Turkish Air Force Academy. This is related to the increase in the demand for proficient English-speaking pilots both in Turkey and around the world. However, this topic needs further investigation for two reasons: First, it is of utmost importance to identify the driving force for aviation students to learn English and their attitudes so that the quality of the Aviation English curricula can be improved, which was identified as an urgent need in Turkish context (Demirdöken, in press). Second, the motivational factors that may impede or foster L2 learning and hence the L2 achievement of aviation students should be explored. The present exploration of the relationship between L2 motivation, achievement motivation, and L2 achievement of aviation students in the Turkish context will, therefore, lead to new insights into aviation students' L2 development process.

Motivation in L2 Learning

Second language (L2) refers to "any languages learned later than in earliest childhood" (Mitchell et al., 2019, p. 1). L2 learning is, therefore, fundamentally different from the acquisition of the native language (L1), and it is influenced by several internal and external factors. Motivation, as one of the determinant factors in L2 attainment (Sung, 2013, p. 19), is defined as the force that drives an individual to accomplish high degrees of achievement and performance and to overcome obstacles to any kind of difference (Tohidi & Jabbari, 2012). Concordantly, Dörnyei (2009) argued that the "self" need to be placed at the center of research into motivation by proposing his seminal L2 Motivational Self-system (L2MSS) framework. It consists of three essential elements: Ideal L2 Self, Ought-to L2 Self, and L2 Learning Experience. While the ideal L2 self refers to individuals' future self-images with regard to L2 and thus serves as a powerful motivator, the ought-to L2 self is associated with attributes that should be possessed to meet the expectations. L2 learning experience "concerns situation-specific motives related to the immediate learning environment and experience" (Dörnyei, 2019, p. 21), such as teachers, peer groups, curricula, and tasks. Considering these components of the L2MSS, it can be hereby argued that motivation should be regarded as the end result of individuals' wishes to reduce the gap between their image of ideal and actual self. L2MSS is also a significant variable in the ESP context since it is particularly interested in the use of target language in different settings (Saqlain et al., 2020). Accordingly, a number of studies have postulated the influence of motivation in several ESP contexts, including Greece (Katsara, 2008), Iran (Pazoki & Alemi, 2019), and the Gulf (Malcolm, 2013). However, the existing literature fails to shed light on the motivational variables influencing the second language learning of aviation students.

L2 Achievement Motivation

Achievement motivation is another significant variable affecting L2 attainment, and it is defined as "motivation relevant to performance on tasks in which standards of excellence are operative" (Wigfield et al., 2007). Therefore, it can result in poor or excellent performance in L2 attainment. Also, individuals are triggered by the desire to either achieve success or avoid failure. For this reason, they are expected to engage in a task if their intrinsic motivation to participate is stronger than the fear of failure. On the other hand, they are likely to avoid participation if their fear of failure outperforms their intrinsic motivation to participate. In terms of task difficulty, while high achievers are likely to prefer challenging tasks, low achievers tend to prefer easier tasks to avoid failure and ensure guaranteed success in this way. To our best knowledge, the one and only research on achievement motivation in the ESP context was conducted by Zhang et al. (2015). They examined the relationship between student nurses' achievement motivation and self-efficacy levels. Although mean self-efficacy scores were low, achievement motivation significantly positively correlated with self-efficacy, and they concluded that necessary measures should be taken to develop the self-efficacy of Chinese student nurses thereby improving their achievement motivation. As can be concluded here, there is scant research on achievement motivation in ESP settings. Therefore, it is the second concern of the present study to investigate how achievement motivation, together with L2MSS, affects L2 achievement in Aviation English courses.

Overall, L2 learning is a long process, and it can be challenging to attain high degrees of motivation in contexts like Turkey, where English is not the common means of communication. Therefore, L2 learners who are not motivated enough may fail to accomplish the goals of the instruction no matter how talented they are or how good the curricula and teaching are (Dörnyei, 2005). Also, they may not achieve high degrees of achievement motivation resulting in failure to attain the target learning outcomes. In this regard, it is hypothesized that there is a significant correlation between ideal L2 self, ought-to L2 self, and achievement motivation and aviation students' L2 achievement in Aviation English courses. The present study seeks answers to the following research questions:

(1) How do L2 motivational variables relate to the English language achievement of aviation students in Turkey?

   a. Is there a statistically significant difference between high- and low-achieving aviation students in terms of L2 achievement?
   b. Is there a statistically significant difference between high- and low-achieving aviation students in terms of L2 motivational variables?
   c. Do L2 motivational variables predict the L2 achievement of aviation students?

(2) What are the attitudes of aviation students in Turkey toward learning Aviation English?

Methodology

Setting and Participants

This study took place between December 2020 and January 2021 at a major state university in Istanbul, Turkey. The aviation English course offered by this institution covered a range of aviation-related topics, including aviation phraseology, radiotelephony communication, Meteorological Aerodrome Reports (METAR), aviation safety, and aeronautical information. Since the nature and purpose of this research required a particular group of aviation students, the researchers applied the following selection criteria:

- students who were in their last year (senior class students)
- students who had not previously taken the Aviation English course
- students who were enrolled in the Aviation English course at the time of the study
- students who had not sat for the American Language Course Placement Test yet
- students who intend to enroll in flight training upon graduation
- students who volunteered for the study

In the end, 111 aviation students who met all the selection criteria were recruited based on the purposive sampling technique.

Instruments, Procedure, and Analysis

Drawing on Ivankova et al. (2006), the present study adopted the mixed-methods sequential explanatory design consisting of collecting and analyzing the quantitative data and then the qualitative data in two consecutive phases. The present study was carried out in two phases, and the data were collected utilizing two sets of questionnaires and semi-structured interviews (Table 1).

Table 1
The instruments, research questions, and purpose

Phase	Research Questions	Data Collection Instrument	Purpose
Phase 1	RQ1: How do achievement motivation, ideal L2 self, and ought-to L2-self affect the L2 achievement of aviation students in Turkey?	L2MSS Questionnaire (Taguchi, Magid, & Papi, 2009). Achievement Motivation Measure (AMM) (Karaman & Smith, 2019)	Identifying the relationship between aviation students' L2 motivational variables and their L2 achievement.
Phase 2	RQ2: What are the attitudes of aviation students toward learning Aviation English?	Semi-structured interviews	Identifying the attitudes of aviation students towards learning Aviation English.

In the first phase, L2 Motivational Self System Questionnaire (Taguchi et al., 2009) and Achievement Motivation Measure (Karaman & Smith, 2019) were instrumented to determine the motivational levels of aviation students. While the former included 20 five-point Likert items, the latter included 13 five-points Likert items. The quantitative data collected through these two questionnaires were analyzed by utilizing SPSS (version 23) software. Regression and correlational analyses were run to explore the relationship between aviation students' L2 motivational variables and achievement. Also, students' scores on the American Language Course Placement Test (ALCPT) were compared to determine their L2 achievement at the end of the semester. The ALCPT was administered at the beginning and at the end of the semester, and the researchers did not have to administer any other testing tool to collect data regarding L2 achievement.

In the second phase, semi-structured interviews were conducted with five aviation students. The purpose was to gain insights into their attitudes towards the Aviation English course. The format of semi-structured interviews was based on a "pre-prepared, elaborate interview schedule/guide" (Dörnyei, 2007, p. 135). It included a specific list of questions to be addressed to each interviewee so that the answers could be compared across different interviewees. The interview questions were generated based on L2 motivational variables, and nine questions were addressed to each interviewee. While six interview questions were related to the L2 motivational self-system, the remainder of the questions were about achievement motivation. Participants' responses were analyzed through the qualitative content analysis method, which is based on the identification of thematic patterns in a data set (Neuendorf & Kumar, 2016). In order to establish intercoder reliability, O'Connor and Joffe (2020) suggest that at least two independent coders are necessary for the coding process. Therefore, both researchers took part in this process independently and coded the qualitative data iteratively by instrumenting open, axial, and selective coding techniques (Boeije, 2010; Flick, 2009). Due to the relatively small volume of qualitative data collected from five aviation students, the researchers did not use any qualitative data analysis software. Rather, in the final stage of the data analysis, the researchers coded ad hoc segments that they determined to be conceptually meaningful, identified the commonalities in responses, and then the emerging codes were grouped into thematic patterns in relation to L2 motivational variables.

Results

The purpose of the present study was twofold. First, the motivational variables were analyzed in relation to the English language achievement of aviation students. Second, the attitudes of aviation students towards the Aviation English course were explored. Overall, four females (3.6%) and 107 males (96.4%) participated in the first phase of the study, and their ages ranged between 20 and 22 ($\bar{x}$ = 21.23), whereas five males (100%) aged between 20 and 22 ($\bar{x}$ = 21.20) made up the participants in the second phase of the study (Table 2).

Table 2

Participants, Demographics, and Interview Data

Phase	Category	Sub-category	N	%	M
Phase 1	Age	20	5	4.5	
		21	75	67.6	
		22	31	27.9	
		Total	111	100	21.23
	Gender	Male	107	96.4	
		Female	4	3.6	
		Total	111	100	
Phase 2	Age	20	1	20.0	
		21	2	40.0	
		22	2	40.0	
		Total	5	100	21.20
	Gender	Male	5	100	
		Female	0	0	
		Total	5	100	

Concerning the differences in the L2 achievement of aviation students, the following conclusions were drawn from the quantitative data analyses. First, the reliability tests showed that Cronbach's Alpha coefficient score of the L2 Motivational Self-System Questionnaire was .764. The same test produced a reliability score of .809 for the Achievement Motivation Measure (Table 3). Therefore, it was concluded that both instruments showed a good degree of reliability, according to Nunnally (1978).

Table 3

Reliability Scores of the Instruments

Instrument	Cronbach's Alpha	N of Items
L2 Motivational Self-System Questionnaire	.809	13
Achievement Motivation Measure	.764	20

The main results of the study were presented based on the research questions respectively in the following titles.

Results regarding RQ1

The first research question sought answers to the motivational variables' relation to aviation students' L2 achievement. Descriptive statistics showed that students' responses varied most on their reported ought-to L2 selves. For instance, aviation students held different opinions on the possible negative future effects of not learning English (ought-to L2 self, Item 7) ($\bar{x}$ = 2.84, *SD* = 1.832), which appeared to be the highest variation among all motivational motives. Regarding achievement motivation, students met on common grounds that while working on a task, they thought of how it would feel when and if the task was successfully completed (Item 13) ($\bar{x}$ = 4.30, *SD* = .880). On the other hand, students' projections for the future mostly included using English (ideal L2 self, Item 8), which received the highest mean score and appeared the most agreed upon construct ($\bar{x}$ = 4.50, *SD* = .796) as well.

In order to analyze the relation of L2 motivational variables to L2 achievement, students were grouped as low and high achievers. The criterion for the grouping was based on students' scores on American Language Course Placement Test (ALCPT). The test-takers must score seventy or higher to be accepted to the pilotage program after graduation, as well as get a fit-to-fly health certificate. The tests are utilized at regular intervals by the institution, and thus the quantitative data regarding participants' performance on ALCPT were retrieved from the institution's database by the researcher for the present study. The test score of seventy on ALCPT, which was conducted at the beginning (Time 1) and end of the academic year (Time 2), was regarded as the cut point to place the participants in one of the two groups, namely low-achievers and high-achievers. In the end, 82 participants were identified as low achievers ($\bar{x}$ = 47.83, *SD* = 7.987) and 29 participants as high achievers ($\bar{x}$ = 75.69, *SD* = 7.431) at Time 1, whereas there were 82 low achievers ($\bar{x}$ = 50.20, *SD* = 9.999) and 29 high achievers ($\bar{x}$ = 78.72, *SD* = 7.240) at Time 2 (Table 4).

Table 4
ALCPT Scores at T1 and T2

		N	Min	Max	Mean	SD
Low achievers	ALCPT Time 1	82	32	65	47.83	7.987
	ALCPT Time 2	82	25	69	50.20	9.999
High achievers	ALCPT Time 1	29	70	100	75.69	7.431
	ALCPT Time 2	29	70	98	78.72	7.240

The Pearson correlation analysis (Table 5) showed that the relationship between achievement motivation levels of aviation students and their ideal L2 selves was positive, moderate in strength (Köklü et al., 2006), and statistically significant (r = .336, n = 111, $p \leq 0.05$, and that the relationship between achievement motivation levels of aviation students and their ought-to L2 selves was negative, weak in strength (Köklü et al., 2006), and statistically significant (r = -.267, n = 111, $p \leq 0.05$).

Table 5
Correlations

Variables	*n*	$\bar{x}$	*SD*	1	2	3
1. Achievement	111	50.22	6.87			
2. Ideal L2 Self	111	36.45	8.43	.336**	-	
3. Ought-to L2 Self	111	21.41	5.95	-.267**	-.020	-

Correlations are significant at the 0.01 level (2-tailed).

Results regarding RQ1a

The purpose of RQ1a was to determine if there existed a statistically significant difference between high- and low-achieving aviation students in terms of L2 achievement. Accordingly, the homogeneity of variances was first tested to determine whether the data was equally distributed among the three constructs being analyzed. Based on the results of Levene's test (Table 6), equality of variances was assumed. Then, the mean scores of both groups on ALCPT at Time 2 were compared to determine how their L2 achievement differed by the end of the semester, and it was found that mean scores had increased for both groups. Besides, paired samples t-test (Table 7) showed that the change in the mean score from Time 1 to Time 2 was positive and statistically significant for both low achievers ($t(81) = 2.727$, $p \leq 0.05$) and high achievers ($t(28) = 4.297$, $p \leq 0.05$). Also, high achievers ($\bar{x} = 3.034$; $SD = 3.803$) improved more than the low achievers ($\bar{x} = 2.366$, $SD = 7.856$) from Time 1 to Time 2.

Table 6
Homogeneity of Variances

	Levene Statistics	df1	df2	Sig.
Achievement Motivation	.900	1	109	.345
Ideal L2 Self	.405	1	109	.526
Ought-to L2 Self	3.040	1	109	.084

Table 7
Paired Samples t-test

		Paired Differences							
					95% Conf. Int. of the Difference				
		$\bar{x}$	*SD*	Std. Error Mean	Lower	Upper	t	df	Sig. (2-tailed)
Low achievers	ALCPT T2	2.366	7.85	.868	4.092	.640	2.727	81	.008
High achievers	ALCPT T2	3.034	3.80	.706	4.481	1.588	4.297	28	.000

Results regarding RQ1b

The purpose of RQ1b was to analyze the differences between the low- and high-achieving groups in terms of L2 motivational variables. Independent samples t-test was run for this purpose (Table 8). While there was a statistically significant difference between the two groups in terms of ideal L2 self ($t(109) = 2.178$, $p \leq 0.05$ and the ought-to L2 self ($t(109) = 2.138$, $p \leq 0.05$), no significant difference was found between two groups in terms of achievement motivation ($t(109) = -.610$, $p > 0.05$). With regard to the ideal L2 self, the high-achieving group had a higher mean score of 39.34 ($SD = 8.09$) compared to the low-achieving group ($\bar{x} = 35.43$, $SD = 8.37$). On the other hand, low achievers had higher mean scores on ought-to L2 self ($\bar{x} = 22.12$, SD = 6,18) compared to high achievers ($\bar{x} = 19.41$, $SD = 4.81$).

Table 8

Independent Samples t-test

Variables	Groups	*N*	$\bar{x}$	*SD*	t	df	*p*
Achievement Motivation	Low	82	49.98	7.10	-.610	109	.543
	High	29	50.89	6.25			
Ideal L2 Self	Low	82	35.43	8.37	2.178	109	.032
	High	29	39.34	8.09			
Ought-to L2 Self	Low	82	22.12	6.18	2.138	109	.035
	High	29	19.41	4.81			

Results regarding RQ1c

Finally, the purpose of RQ1c was to determine if achievement motivation and L2MSS could significantly predict ESP students' L2 achievement. Results of the multiple regression analysis (Table 9) showed that L2 motivational variables could explain 15% of the total variance ($r^2 = 0.15$; $F(3,107) = 6.533$; $p \leq 0.05$), which proved that L2 motivational variables were moderate in strength and statistically significant predictors of L2 achievement of aviation students. It was also found that the ideal L2 self ($\beta = .532$; $p \leq 0.05$) and the ought-to L2 self ($\beta = -.703$; $p \leq 0.05$) significantly predicted L2 achievement, whereas achievement motivation ($\beta = -.179$; $p \leq 0.05$) did not significantly predict it.

Table 9

Multiple Regression Analysis

	Unstandardized Coefficients			
Variables	β	Std. Error	t	Sig.
(Constant)	61.010	12.400	4.920	.000
Achievement Motivation	-.179	.210	-.852	.396
Ideal L2 Self	.532	.165	3.226	.002
Ought-to L2 Self	-.703	.229	-3.076	.003

$r = .393$; $r^2 = .155$; $p = 0.000$

Results regarding RQ2

The purpose of the second research question was to gain insights into aviation students' attitudes toward the Aviation English course. Data were collected from five participants between January 5 and 11, and interviews lasted around 12 minutes (Table 10). Data were analyzed qualitatively on MAXQDA software, and three main themes were identified: personal future goals, professional concerns, and personal attachment.

Table 10
Interview Data

Participants	Date	Duration
P1	05.01.2021	13 minutes
P2	05.01.2021	11 minutes
P3	08.01.2021	14 minutes
P4	11.01.2021	15 minutes
P5	11.01.2021	11 minutes

Personal future goals

A key theme identified in this study was participants' *personal future goals*, which were mostly reported concerning the ideal L2 self. P1, for instance, expressed his motivation to learn Aviation English (AE) with a direct emphasis on his future personal goal: "*I can define it [referring to AE] as a must-have for me. I am here to be a pilot so learning English will help me a lot in my work environment."* In another instance, P5 raised the same issue: "*I want to be a pilot in the future so learning ATC communication, aircraft parts and other aviation terms will contribute a lot to me, and I will be ready for my profession.*" It seemed that participants were motivated to learn Aviation English, which played a significant role in their future plans. It was illustrated by P4 as follows: "*It will help me a lot in my future profession. I should represent my country well and show everyone that we are leading. This is only possible with good control of Aviation English.*" The participants also had clear projections for the future, and their ideal L2 selves were shaped by examples from real life. For instance, P5 imagined himself "*...presenting a brand-new aircraft to other people*" whereas P4 stated, "*I can imagine myself talking to ATC during a flight*". Similarly, "*…talking to other people easily and having a conversation with them about a joint flight operation*" was how P1 imagined himself in the future.

Professional Concerns

It was also found that aviation students' attitudes toward learning Aviation English were shaped by their concerns regarding the target workplace. However, these concerns were mostly related to other people's expectations. Therefore, this theme was associated with the ought-to L2 self, and it was called *professional concerns*. Participants reported that there was high expectation from them, and there would be some negative outcomes if they could not meet these expectations. P1, for instance, mentioned one of these negative outcomes in the workplace: "*They expect me to learn Aviation English to serve best for my own institution… I will have a lot of difficulties and I will fall behind my colleagues. However, I would be a reputable and* *trustworthy person if I learn.*" The way other people would think about him was a significant concern. In a similar vein, P2 stated that "*There is a huge expectation. It would definitely be a good thing for me to learn Aviation English. Also, I would have an advantage over my colleagues in our base.*" P2 attributed learning Aviation English to taking advantage in the workplace. On the other hand, P4 did not report any consequence of not learning Aviation English in the way other people would think about him. He said: "*It does not make much difference. If I learn it, I do it for myself; if I do not learn it, it will not change other people's opinion about me.*" Rather, he was more concerned about its negative consequences for his career, reporting:

> It would be a huge waste of time for me if I could not learn Aviation English until my graduation. I cannot spend more time learning Aviation English because I should be ready to learn how to fly when the time comes.

P3 also reported that in the event of not learning Aviation English:

> I will fail because they only expect me to learn aviation terminology. As I do not have any experience right now, I cannot think of any benefits of learning Aviation English to change the opinions other people hold about me.

Aviation students were aware of the fact that failure to learn AE would have negative consequences for them, but they held conflicting ideas on how it would influence others' opinions about them.

Personal Attachment

Finally, the data analysis showed that participants were personally attached to their profession, and it served as the source of motivation for them. Therefore, this theme was identified as personal attachment, and it was associated with achievement motivation. In this regard, P2 was content with his current effort to learn Aviation English as he reported, "*It is very meaningful for me because I want to be in the shoes of those I see in videos.*" The same contentedness was reflected by P2 as well. He said: "*The examples I see motivate me a lot, so I am quite motivated.*" On the other hand, P4 was more pragmatic, highlighting his interest in only what he will make use of in the future: "*I am focused on what I will need in the future. These include aviation terminology and ATC communication. Therefore, I am doing my best.*" Also, P1 held a relatively lower degree of motivation as he simply stated that "*I believe it is intermediate level.*" Their attachment to their profession was a significant contributory factor to what kept aviation students motivated throughout their L2 learning experience. For instance, P4 mentioned enjoying his AE lessons: "*It is quite enjoyable. I always imagine myself as a pilot, so I have a lot of fun learning aircraft parts and how to communicate over the radio. These all keep me motivated.*" Besides having fun, aviation students' attachment to their future profession was associated with their area of interest and plans. In this regard, P1 reported that "*Constantly, I think of my plans because they are all related to having proficiency in AE. That is how I keep motivated all the time.*" His future projections required being a proficient speaker of AE, and thus, he had a higher degree of achievement motivation.

Discussion and Conclusion

The present study identified the influence of L2 motivational variables on academic achievement. Also, it uncovered student pilots' attitudes towards the Aviation English course. In this sense, it made a substantial contribution to ESP motivation research by focusing on this uncharted context.

Regarding Dörnyei's (2009) L2MSS model, it was conceptualized that the ideal L2 self and ought-to L2 self significantly correlate with aviation students' L2 achievement in Aviation English courses. As hypothesized, the former significantly positively correlated with L2 achievement. Although this finding was in line with the study of Lamb (2012) conducted in Indonesia and Dörnyei and Chan (2013) in China, it conflicted with the findings of Moskovsky et al. (2016), who carried out research in Saudi Arabia and Subekti's (2018) finding in Indonesia. The most reasonable explanation for such conflicting results in different contexts can be attributed to the fact that “reported motivation does not always have behavioral consequences" (Moskovsky et al., 2016) in all settings. The participants in the present study might have already developed clear projections for the future, and this might have positively affected their perceived L2 ideal selves, thus contributing to their L2 achievement. This is evidenced by the fact that this issue was clearly highlighted during the interviews as a source of motivation to learn aviation English, and it also appeared as the most agreed-upon construct in our setting, where students are specifically trained to be a pilot. However, the participants in Subekti's (2018) study were taking an EAP course designed to equip students with reading skills for a better comprehension of scientific articles and journals. In this sense, these EAP students might not have developed an explicit reason to learn English as part of their profession. Similarly, the participants in Moskovsky and others' (2016) study were EFL students with a low ideal L2 self, resulting in a negative correlation with L2 achievement. Therefore, it can be hereby argued that developing a clear-cut projection for the future may have positive implications for students in the context of aviation.

As for the second construct in Dörnyei's (2009) L2MSS model, the ought-to L2 self significantly negatively correlated with aviation students' L2 achievement in Aviation English courses. That is, the higher the ought-to L2 self was, the lower the aviation students tended to perform. However, this finding was not in accordance with Moskovsky et al. (2016), Papi and Abdollahzadeh (2012), and Dörnyei and Chan (2013) in that no significant relationship was found between ought-to L2 self and L2 achievement in those studies. This could be partially attributed to exam-oriented settings like Saudi Arabia (Moskovsky et al., 2016) and Iran (Papi & Abdollahzadeh, 2012), where the score on nationwide proficiency exams is the main criterion for students. Therefore, learning English is mostly mediated by societal factors in those settings. At this point, we must differentiate between such settings and the target setting of the present study. Although the participants in this study are required to score seventy or more on ALCPT to be accepted to the pilotage program in the USA, they sit for that test several times before they graduate. Therefore, they are likely to be influenced to a limited degree by their overall performance on such tests compared to those in the Iranian or Saudi Arabian contexts. Another reason could be the fact that ‘externally sourced self-images’ fail to serve as the triggering power that can make a difference in actual learner behavior (Dörnyei & Chan, 2013). Consequently, learners do not engage in any meaningful activity to learn the language, and thus weaker links appear between ought-to L2 self and L2 achievement. As it turned out in the present study, aviation might be a distinct context in this regard, yet future research in a similar context is essential to further investigate the statistically significant yet negative relationship between ought-to L2 self and L2 achievement.

With regard to the second motivational variable of the present study, no significant correlation was found between achievement motivation and L2 achievement. The existing literature showed conflicting results with the findings of the present study. While Emmanuel et al. (2014) concluded in line with our study, Tella (2007) and Sikhwari (2014) concluded on the contrary, presenting a significant and positive impact of achievement motivation on L2 achievement. It is possible that Covid-19 might have had detrimental effects on the participants of the present study. This is similar to the problem originally encountered in a recent study by Zaccoletti et al. (2020) who revealed that there was a significant decrease in Italian and Portuguese students' motivation with the outbreak of the Covid-19 pandemic. In addition, Dörnyei (2005) suggests that language learning is a long process and thus unmotivated students may fail to attain the learning outcomes regardless of their talent, and the quality of curricula and teaching. It may be the case therefore that student pilots' achievement motivation did not correlate with their achievement.

Another point we focused on was the predictive power of the ideal L2 self, ought-to L2 self, and achievement motivation on aviation students' L2 achievement in Aviation English courses. We found out that although Dörnyei's (2009) ideal L2-self and ought-to L2-self constructs could significantly predict aviation students' L2 achievement in Aviation English courses, achievement motivation failed to predict it. Our finding regarding the predictive power of these L2 motivational constructs is significant in that the existing literature fails to provide an account of how L2 motivational variables affect L2 achievement in a goal-oriented ESP teaching setting. Therefore, we have hereby provided an important insight into future ESP pedagogy. Apart from the quantitative evidence regarding the predictive role of the ideal L2 self on L2 achievement, it is also evidenced in semi-structured interviews that aviation students' ideal L2 self-images serve as a powerful motivator to learn Aviation English. For instance, P1 and P5 had explicitly reported that their motivation to learn aviation English was instrumented by their desire to be a pilot. Similarly, P5 depicted himself talking with pilots from other nations during a joint flight operation. However, P4 had reported that there was a big expectation and thus he focused on what he would need in the future. Therefore, developing an ideal L2 self and staying motivated towards their future goals can only be ensured if ESP pedagogy offers hands-on experience and a vast amount of opportunity for aviation students to speak as themselves in the target setting. Also, ESP teachers should design unique course materials presenting real-life use of the target language. Thus, aviation students can feel more attached to their future profession, and they can develop a better L2 ideal self-image for learning Aviation English.

The present study also offered insights into aviation students' attitudes toward learning Aviation English with regard to the ideal L2 self, ought-to L2 self, and achievement motivation. As stated earlier in this paper L2 learning is a long process and thus, it is crucial to be able to stay motivated. Otherwise, aviation students are likely to fail to attain proficiency in the target language. As for our sampling, learning Aviation English was fun and they had a constant interest in aviation, which was reflected in interview responses. It can be argued that these students' previous L2 learning experience in Turkey mostly included traditional teacher-centered approaches with little or no focus on communicative competence. Also, the participants in the present study would no longer have to take any nationwide proficiency exam to practice their profession; rather they would sit on the American Language Course Placement Test with which they were already familiar. Therefore, their previous L2 learning experience might have impeded the development of a high degree of ideal L2 self. However, experiencing a student-centered approach with a focus on communicative competence might have paved the way for aviation students for faster development of such high degrees of ideal L2 self. This was in line with the findings of Taguchi et al. (2009) in that classroom experience was important for Japanese and Iranian students. Similar to our sampling students, attitudes to learning English played a significant role for Japanese and Iranian students. Contrastively, enjoyment did not play a decisive role in Chinese students' overall motivation in the same study. On the other hand, Turkish aviation students' enjoyment seemed to have contributed a lot to their overall motivation mostly due to contextual differences.

All things considered, Dörnyei's (2005, 2009) L2MSS was found to account for L2 achievement in the Aviation English teaching/learning setting. However, it was not possible to conclude that achievement motivation resulted in poor or excellent performance in L2 attainment in the same setting. Also, aviation students' attitudes toward learning Aviation English were mostly impacted by the attributes they would like to have in the future. Although expectations of other people influenced the same construct, these expectations were not as powerful as internalized instrumental motives.

Implications

The present study was exploratory in nature and researchers aimed to explore the relationship between L2 motivational variables and aviation students' L2 achievement. In light of our findings, the following issues should be considered by ESP practitioners. First, multiple regression analysis and responses to semi-structured interview questions showed that developing a clear self-image for the future is likely to bring success to aviation students. Second, aviation students differ from ESL/EFL students in that the former usually have pre-set future projections and these projections serve as a powerful source of motivation to learn the target language. Therefore, ESP courses should be designed in such a way that they can sustain aviation students' interest and address their communicative needs in the target work environments. These may include authentic materials like interviews with senior pilots, recordings of real-life ATC radiotelephony communications, and field trips. Aviation students can, in turn, get the most out of Aviation English courses, and thus the demand of the aviation industry for proficient English-speaking pilots can be met in the future.

Limitations

The present study was conducted in a non-English-medium instruction (EMI) setting. Therefore, motivational variables affecting L2 achievement may yield different outcomes in EMI ESP settings. Also, only three motivational variables, namely ideal L2-self, ought-to L2-self, and achievement motivation, were considered. However, motivation is complex and thus, other factors that were not considered in the scope of this study might also influence motivation.

Another limitation of the present study was related to the measure that was used to determine achievement. Since the participants were assessed based on their score on the ALCPT, the measurement was limited to the scope of this language test, which assessed test takers' listening and reading comprehension skills as well as their vocabulary and grammar knowledge. Therefore, the standard measurement in this test lacked speaking and writing skills. It should also be noted that data were collected in the middle of a global pandemic and the effects of this pandemic are unknown. Therefore, this should be considered when generalizing the findings of this study. Last but not least, the present study was carried out in Turkey where English was not the native language of the participants. Therefore, student pilots' motivation to learn English as a second language in Turkey may not be generalized to other contexts.

Funding: The authors received no financial support for the research, authorship, and/or publication of this article.

Declaration of Interest: We hereby declare that the researchers in the present study had no competing interests.

References

Alfehaid, A., & Alkhatib, N. (2020). ESP: A Real-life Perspective. *Asian ESP Journal*, *16*(2.1), 159-175.

Boeije, H. (2010). *Analysis in Qualitative Research.* SAGE.

Dörnyei, Z. (2005). *The psychology of the language learner: Individual differences in second language acquisition.* Erlbaum.

Dörnyei, Z. (2007). *Research Methods in Applied Linguistics.* Oxford University Press.

Dörnyei, Z. (2009). The L2 motivational self-system. In Z. Dörnyei, & E. Ushioda (Eds.), *Motivation, language identity and the L2 self* (pp. 9-42). Multilingual Matters.

Dörnyei, Z. (2019). Towards a better understanding of the L2 Learning Experience, the Cinderella of the L2 Motivational Self System. *Studies in Second Language Learning and Teaching, 9*(1), 19-30.

Dörnyei, Z. & Chan, L. (2013). Motivation and Vision: An Analysis of Future L2 Self Images, Sensory Styles, and Imagery Capacity Across Two Target Languages. *Language Learning, 63*(3), 437-462. https://doi.org/10.1111/lang.12005

Emmanuel, A. O., Adom, E. A., Josephine, B., & Solomon, F. K. (2014). Achievement Motivation, Academic Self-Concept, and Academic Achievement among High School Students. *European Journal of Research and Reflection in Educational Sciences, 2*(2), 24-37.

Flick, U. (2009). *An introduction to qualitative research.* SAGE.

ICAO. (2004). *Manual on the implementation of ICAO language proficiency requirements.* Montreal, Quebec, Canada: ICAO Doc 9835 AN/453.

ICAO. (2020). *Safety Report.* Montreal: International Civil Aviation Organization.

Ivankova, N. V., Creswell, J. W., & Stick, S. L. (2006). Using Mixed-Methods Sequential Explanatory Design: From Theory to Practice. *Field Methods, 18*(1), 3-20. https://doi.org/10.1177/1525822X05282260

Karaman, M. A., & Smith, R. (2019). Turkish Adaptation of Achievement Motivation Measure. *International Journal of Progressive Education, 15*(5), 185-197.

Katsara, O. (2008). Aspects of motivation within the context of an ESP course. *English for Specific Purposes, 7*(3), 1-26.

Köklü, N., Büyüköztürk, S., & Coklu, Ö. (2006). *Sosyal Bilimler İçin İstatistik.* Pegem.

Lamb, M. (2012). A self-system perspective of young adolescents' motivation to learn English in urban and rural setting. *Language Learning, 62*(4), 997–1023.

Leech, N. L., Barrett, K. C., & Morgan, G. A. (2008). *SPSS for intermediate statistics: Use and interpretation (3rd ed.)*. Lawrence Erlbaum Associates Publishers.

Malcolm, D. (2013). Motivational challenges for Gulf Arab students studying medicine in English. In Ushioda, E. (Ed.), *International Perspectives on Motivation* (pp. 98-116). Palgrave Macmillan.

Maxwell, J. (1997). Designing a qualitative study. In Bickman, L., & Rog, D. J. (Eds.), *Handbook of Applied Social Research Methods* (pp. 69-100). SAGE.

Mede, E., Koparan, N., & Atay, D. (2018). Perceptions of students, teachers, and graduates about civil aviation cabin services ESP program: An exploratory study in Turkey. In Kırkgöz, Y., & Dikilitas, K. (Eds.), *Key Issues in English for Specific Purposes in Higher Education* (pp. 157-174). Springer.

Moskovsky, C., Racheva, S., Assulaimani, T., & Harkins, J. (2016). The L2 motivational self-system and L2 achievement: A study of Saudi EFL learners. *The Modern Language Journal, 100*, 1–14.

Neuendorf, K. A., & Kumar, A. (2016). *Content Analysis. The International Encyclopedia of Political Communication.* John Wiley & Sons. https://doi.org/10.1002/9781118541555.wbiepc065

Nunnally, J. C. (1978). *Psychometric theory (2nd ed.).* McGraw-Hill.

O'Connor, C., & Joffe, H. (2020). Intercoder Reliability in Qualitative Research: Debates and Practical Guidelines. *International Journal of Qualitative Methods*, *19*:1-13. https://doi:10.1177/1609406919899220

Papi, M., & Abdollahzadeh, E. (2012). Teacher Motivational Practice, Student Motivation, and Possible L2 Selves: An Examination in the Iranian EFL Context. *Language Learning, 62*(2), 571-594.

Pazoki, S. J., & Alemi, M. (2019). Engineering Students' Motivation to Learn Technical English in ESP Courses: Investigating Iranian Teachers' and Students' Perceptions. *RELC, 51*(2), 212-226.

Saqlain, N., Shafqat, A., & Hassan, A. (2020). Perception Analysis of English Language Teachers about Use of Contextualized Text for Teaching ESP. *Asian ESP Journal, 16*(5.1), 275-297.

Sikhwari, T. D. (2014) A Study of the Relationship between Motivation, Self-concept and Academic Achievement of Students at a University in Limpopo Province, South Africa.

International Journal of Educational Sciences, *6*(1), 19-25. https://doi:10.1080/09751122.2014.11890113

Subekti, A. S. (2018). L2 Motivational Self System and L2 achievement: A study of Indonesian EAP learners. *Indonesian Journal of Applied Linguistics, 8*(1), 57-67.

Taguchi, T., Magid, M., & Papi, M. (2009). The L2 motivational self-system among Japanese, Chinese, and Iranian learners of English: A comparative study. In Dörnyei, Z., & Ushioda, E. (Eds.), *Motivation, Language identity, and the L2 Self* (pp. 66-77). Multilingual Matters.

Tella, A. (2007). The impact of motivation on student's academic achievement and learning outcomes in mathematics among secondary school students in Nigeria. *Eurasia Journal of Mathematics & Technology Education, 3*(2), 149-156.

Tohidi, H., & Jabbari, M. M. (2012). The effects of motivation in education. *Procedia, Social and Behavioral Sciences, 31,* 820-824.

Trippe, J., & Baese-Berk, M. (2019). A Prosodic Profile of American Aviation English. *English for Specific Purposes, 53*, 30-46.

Wigfield, A., Eccles, J. S., Schiefele, U., Roeser, R. W., & Davis-Kean, P. (2007). Development of Achievement Motivation. In Damon, W., Lerner, R. M., & Eisenberg, N. (Eds.), *Handbook of Child Psychology* (pp. 933-1002). John Wiley & Sons.

Zaccoletti, S., Camacho, A., Correia, N., Aguiar, C., Mason, L., Alves, R. A., & Daniel, J. R. (2020). Parents' Perceptions of Student Academic Motivation During the COVID-19 Lockdown: A Cross-Country Comparison. *Frontiers in Psychology, 11*:592670. https://doi:10.3389/fpsyg.2020.592670

Zhang, Z., Zhang, C., Zhang, X., Liu, X., Zhang, H., Wang, J., & Liu, S. (2015). Relationship between self-efficacy beliefs and achievement motivation in student nurses. *Chinese Nursing Research, 2*, 67-70.

Collegiate Aviation Review International

Volume 40 | Issue 2 Peer Reviewed Article #5

11-5-2022

An Exploration of the Relationship Between Flight Simulator Performance and Achievement of Solo Flight Among Australian Aviation Students

Benjamin Michael Quast
University of South Australia

Harry James Clarke
University of South Australia

Shawn Lee
University of South Australia

Kaartik Walia
University of South Australia

Steven Leib
Central Queensland University

University flight training programs are becoming an increasingly important avenue for developing ab initio pilots, yet training programs suffer high attrition rates. Flight simulators are commonly used by university flight schools as a training aid, and the purpose of this research is to understand if student performance using a Personal Computer-Based Aviation Training Device (PCATD) is a relevant predictor of student success as measured by the achievement of flying solo in university flight training. To investigate this, 195 students at an Australian university from 2018 to 2021 were subject to comprehensive flight simulator instruction via a PCATD prior to flight training, with simulator performance correlated to flight training success. This sample was split into international and domestic students, with the PCATD performance of each group correlated to the achievement of the first solo and the number of flight hours to the first solo, respectively. Results suggested that international students who achieved the first solo had better simulator performance on average than those who did not. However, a statistically significant relationship was unable to be observed between flight simulator performance and flight time to achieve solo flight amongst domestic students.

Recommended Citation:
Quast, B. M., Clarke, H. J., Lee, S., Walia, K. & Leib, S. (2022) An exploration of the relationship between flight simulator performance and achievement of solo flight among Australian aviation students. *Collegiate Aviation Review International*, 40(2), 67-78. Retrieved from http://ojs.library.okstate.edu/osu/index.php/CARI/article/view/8741/8432

Introduction

The decision to embark on flight training represents a considerable investment of time and money from the student and flight training institutions alike (Lutte & Mills, 2019). This resource demand, coupled with the competitive nature of the pilot profession, limited training positions (Lutte & Mills, 2019), and trends of fewer students completing university pilot training (United States Government Accountability Office, 2014), emphasizes the need for effective candidate selection. To assist with identifying suitable prospective candidates, an array of testing methods are employed by flight training institutions to varying degrees of success. The imperfect nature of pilot selection is highlighted by flight university attrition rates, with reports of between 30% (Bjerke & Healy, 2010) and as high as 70% (Peppler, 2011, as cited in Emery, 2011). Flight simulators are commonly used by university flight schools (Goetz et al., 2015), often as a preparatory aid prior to offering official flight training. The training benefits of simulators are widely accepted in aviation, but limited information exists on simulators as a tool to assess the likelihood of student success in university flight training. The purpose of this study was to better understand the relationship between simulator performance and student success in university flight training.

Literature Review

Pilot Selection

Historically, the pilot selection process has developed alongside the aviation industry. An understanding of the favorable personal attributes of pilots has been sought by training institutions as early as World War I and has continued ever since, with pilots enduring as one of the most tested modern professions (Martinussen & Hunter, 2017). Current global aviation training trends are seeing tertiary education playing an increasingly important role in training ab initio pilots, particularly in the civil sector (Lutte & Mills, 2019).

The conventional performance metric to predict student performance at university has been past academic success (Jones, 2013; Zierke, 2014). This aspect is consistent with tertiary flight training, where the grade point average (GPA) has been found to be a valid indicator of future training success in some aspects of flight training (Jones, 2013; McFarland, 2017). Jones' (2013) study of 264 participants found a statistically significant relationship between GPA with the length of time to complete an instrument rating and instrument rating flight test pass rate. A higher GPA was shown to reduce the time taken to complete the rating and the number of flight test attempts required to pass. However, GPA was not found to relate to either length or time to complete or the flight test pass rate for a private rating, nor the number of flight hours to complete either a private or instrument rating.

Whilst GPA makes use of available academic data, some research has been conducted using supplementary testing measures seeking to measure specific attributes. One such attribute is perceptual speed, defined as the “speed of processing information and the ability to focus attention” (Mekhail et al., 2010, p. 106), and is measured using a table reading test or table speed test (Mekhail et al., 2010; McFarland, 2017). A study of 116 participants at a major American university found that perceptual speed could be used to predict time to solo and flight time to achieve a private pilot certificate (Mekhail et al., 2010). However, this finding failed to be replicated by a similar study of 69 participants, where no statistically significant relationship was found between these variables (McFarland, 2017).

Flight Simulators and University Flight Training

Another potential avenue to predict student performance in university flight training is using work sample tests, which are among the best predictors of future performance in flight (Martinussen & Hunter, 2017). A work sample test endeavors to recreate an environment that replicates on-the-job performance; for pilots, a work sample test is typically administered by a simulator (Hoermann & Goerke, 2014). Simulators have been a part of aviation since the Link “blue box” trainer was developed in 1929 (Ennis, 2009). NASA’s Apollo missions also accelerated advancements in simulator technology, leading to the current range of sophisticated systems (Page, 2000), with these high-fidelity simulators typically the domain of military and airlines (McDermott, 2005). The advancement of computer technology through the 1990s lead to the rise of Personal Computer-Based Aviation Training Devices (PCATDs) (Martinussen & Hunter, 2017; McDermott, 2005), which function as rudimentary versions of the high-fidelity simulators (Ennis, 2009). Extensive research on PCATDs, now generally referred to as Flight Training Devices (FTDs), has found that despite the lower fidelity, they are still effective training aid (Martinussen & Hunter, 2017). One method to evaluate PCATD effectiveness was using a concept of training effectiveness ratio (TER). TER is defined as “the degree to which hours in the simulator replace hours in the aircraft” (Roscoe, 1980, as cited in Martinussen & Hunter, 2017). For example, a TER of 1.0 means every hour in the simulator saves an hour of flight time. Similarly, a TER of 0.5 means every hour in the simulator saves a half hour of flight time.

The training and cost-effectiveness of PCATDs have seen wide-scale adoption by tertiary institutions to supplement flight training (Goetz et al., 2015). In a university flight training setting, PCATDs have been shown to achieve positive TERs ranging from 0.12 up to 0.28 (Taylor et al., 1999, as cited in Martinussen & Hunter, 2017). PCATD training was found to decrease the flight hours prior to the first solo and landings prior to the first solo in some ab initio students prior to commencing flight training (Vaden et al., 1998). Also, a study of 14 participants from the beginning of university flight training found that those using the PCATD performed better than the control group on average, but the difference was not statistically significant (Olson, 2002).

Research Questions and Hypotheses

The aim of this research was to further explore the relationship between simulator test grade (STG) performance and flight training proficiency through the lens of two research questions:

> Research Question 1 (RQ1): How does an international university aviation student's STG relate to the likelihood of the eventual achievement of solo flight?
>
> Research Question 2 (RQ2): How does a domestic university aviation student's STG relate to the number of flight hours accrued prior to achieving solo flight?

These research questions are aligned with the following hypotheses:

> The null hypothesis for RQ1 ($H1_0$): There is no significant difference in STG between students who later achieved solo flight and those who did not.
>
> The alternative hypothesis for RQ1 ($H1_A$): There is a significant difference in STG between students who later achieved solo flight and those who did not.
>
> The null hypothesis for RQ2 ($H2_0$): There is no correlation between STG between students and the number of flight hours accrued prior to achieving solo flight.
>
> The alternative hypothesis for RQ2 ($H2_A$): There is a correlation between STG between students and the number of flight hours accrued prior to achieving solo flight.

Methodology

Design and Sample

This study was quantitative in nature and utilized a sample of university students at a major Australian university to represent tertiary aviation education students in Australia. Eligible participants must have completed a 13-week first-year university simulator subject between 2018 and 2020 and received an STG, as well as undertaken flight training up to at least the point of the first solo.

To answer the research questions, the sample was subdivided into two groups, international and domestic flight training students, which was necessary due to differences in the nature of flight training pathways at the university. After completing the first-year simulator subject, international students were automatically permitted to enroll in a basic flight training course (allowing them to train up to and possibly including solo flight) approximately four to six months after completing flight training. Conversely, domestic students completed a rigorous application process to commence Commercial Pilot Licence training approximately one year after receiving their STG. While flight training pathways differed for international and domestic students administratively, the training content remained the same, allowing both groups the opportunity to progress to solo flight.

Data Collection

This protocol was approved by the university's Human Research Ethics Committee. Data were extracted from existing university student records, collated, and deidentified prior to analysis. No personal information was recorded in the data set, which included only STG scores and solo flight outcomes. STG data was obtained from the first-year simulator subject grade books between 2018 and 2020, corresponding to three full cohort cycles, and was applicable to both sample groups. The flight simulator used to administer training and assess the STG was a desktop PCATD that replicates a Cessna C-172 aircraft using Lockheed Martin Prepar3d software. The PCATD hardware comprises a visual display coupled with force feedback controls consisting of a desktop yoke and throttle quadrant with floor-mounted rudder pedals. The STG was a numerical value out of 20 and was based on a student's manual handling and procedural compliance while flying a circuit from take-off to landing (or go-around).

Data related to solo flying used to answer RQ1 and RQ2 differed due to how the university recorded progress for the two groups. For international students, only a fixed training block of flight hours was available, and training was concluded once the hours had elapsed, whether the solo flight was achieved or not. For domestic students, an hourly limit was not enforced. Hence for the purpose of this study, training success was defined by the number of flight hours required to achieve the first solo. This impacted the data that was collected: for international students, only data on whether a solo flight was achieved as part of their undergraduate flight training subject was recorded. For domestic students undertaking Commercial Pilot License training, university records included the number of flight hours students required to achieve solo flight, which was extracted for RQ2.

Method

For RQ1, an analysis of variance (ANOVA) was performed on the international student data to determine any difference in the mean STG between the students who achieved solo and those who did not. Means and standard deviations for STG scores were compared between students who were ultimately able to achieve solo flight and those who were not. For RQ2, single linear regression was used to explore any relationship between domestic students' STG and corresponding flight time to first solo as measured by flight hours.

Results

Data from 195 students were included in the study, covering124 international students and 71 domestic students. Sample descriptive statistics relating to STG scores and solo performance are displayed in Table 1.

Table 1
Sample descriptive statistics

	International students (N = 124)		Domestic students (N = 71)	
Variable	N	%	N	%
Solo status				
Yes	46	37.1	71	100
No[a]	78	62.9	0	0
	Mean	SD	Mean	SD
STG – yes solo	15.78	2.48	15.15	2.54
STG – no solo[a]	13.65	2.49	-	-
Time to solo[b] (hr)	-	-	20.10	5.86

[a] Domestic students must have flown solo to be considered
[b] Data not considered for international students as outside the scope of this study

Research Question 1

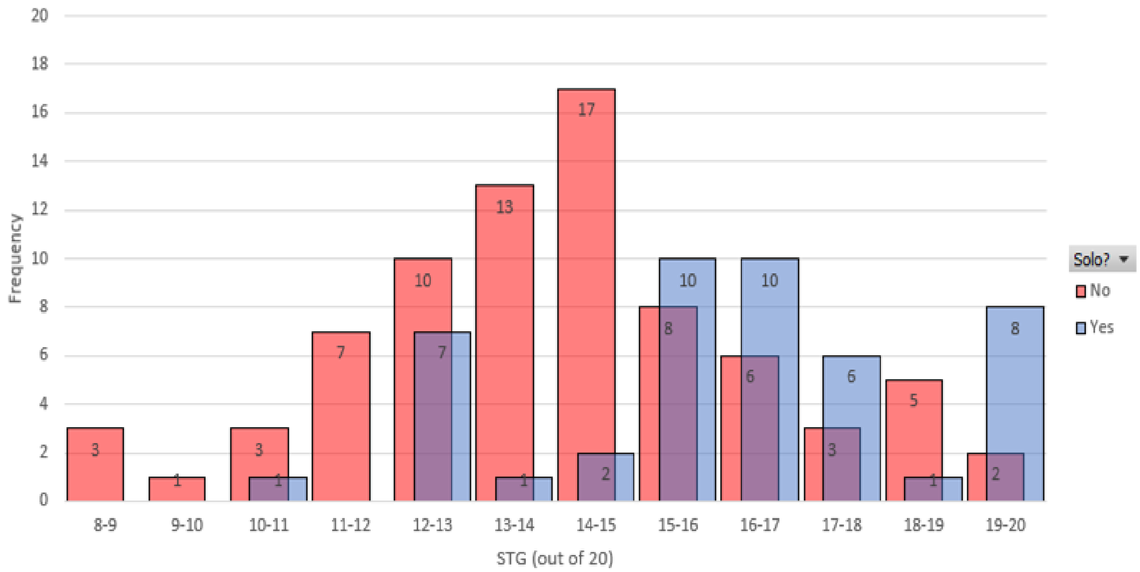
Of the 124 international students in the data set, 78 were unable to achieve solo flight compared to 46 who were. Within these groups, Shapiro-Wilk tests of normality yielded p-values of .133 and .113, respectively, suggesting normality. Additionally, standard errors of 2.49 and 2.48, respectively, yielded a Levene statistic of .000, corresponding to p<.989, suggesting equal variances. As shown in Table 2, international students who did fly solo averaged STG scores of 15.78/20, compared to 13.65/20 for those who did not. This variance corresponded to an F-value of 21.19 and a p-value of <.001.

Table 2
International Student ANOVA Results

	International students (N = 124)	
Variable	M	Var
STG – yes solo	15.78	6.15
STG – no solo	13.65	6.21
F value	21.19	
P value	1.03×10^{-5}	
F crit	3.92	

A histogram for comparative visualization of international student data is displayed in Figure 1.

Figure 1
Histogram of STG vs. Solo flight status for international students



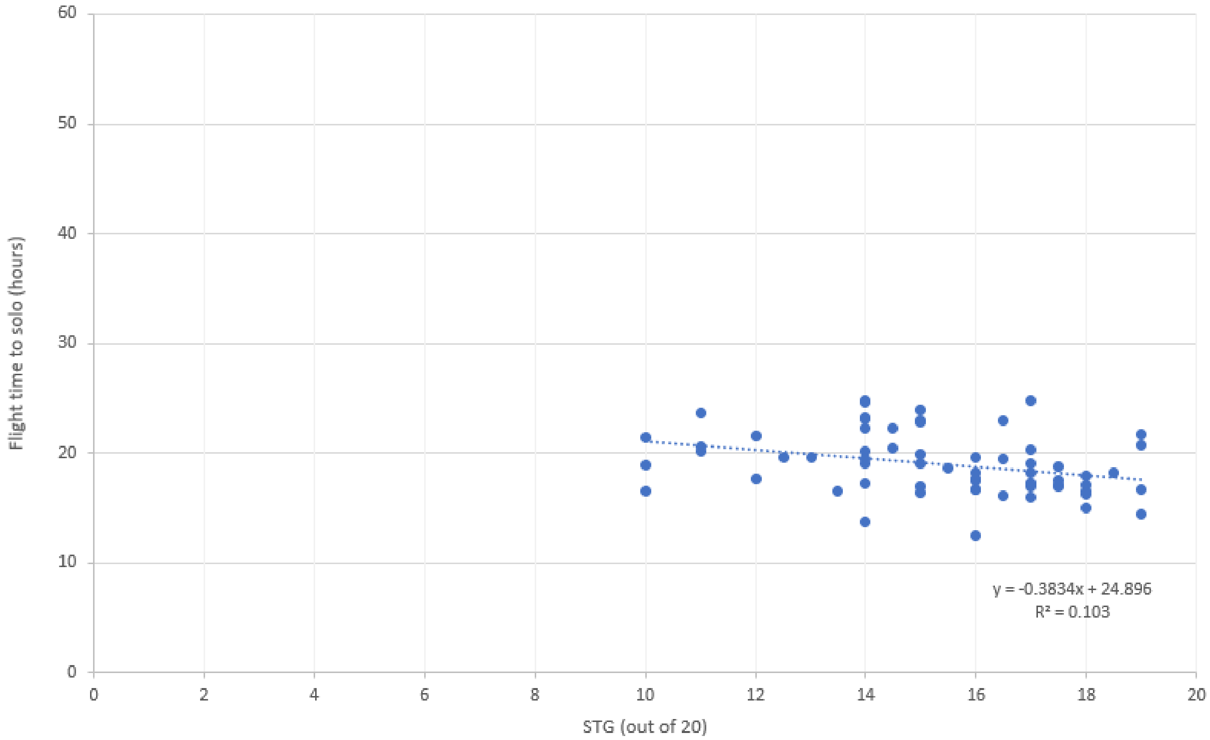
Research Question 2

A linear regression was performed on the domestic student data to explore the relationship between STG and flight time to solo. The descriptive statistics and regression results are displayed in Table 3 and visually displayed via scatterplot (Figure 2). The sample was subsequently analyzed for outliers. Values residing outside 1.5 times the upper or lower interquartile range (IQR) for either variable were deemed outliers. A linear regression was performed with the outliers removed. The subsequent descriptive statistics are displayed in Table 3 and visually displayed via scatterplot (Figure 2).

Table 3
Domestic student descriptive statistics and regression results

	Domestic students (N = 71)		Domestic students – outliers removed) (N = 64)	
Variable	Time to solo	STG	Time to solo	STG
Mean	20.10	15.15	19.00	15.37
Median	18.90	15.50	18.65	16.00
SD	5.86	2.54	2.81	2.35
IQR	4.70	3.00	3.68	3.00
R^2	0.0464		0.103	
R	-0.215		-0.321	

Figure 2
Plot of STG vs. flight time to solo for domestic students – no outliers (N = 64)



Discussion

Strong findings from research RQ1 show that international students with better simulator performance are more likely to achieve solo with a fixed-time training syllabus. With successful students achieving a mean STG of 15.78 and unsuccessful students achieving a mean STG of 13.65, the analysis found a statistically significant difference in simulator performance based on whether students were later able to achieve solo flight. The analysis demonstrates a statistically significant correlation (P-value < 0.05) between STG and attainment of the first solo for international students, therefore rejecting the null hypothesis.

For RQ 2, results of the linear regression analysis indicated a weak negative linear relationship between STG and flight time to solo in domestic students ($R = -0.215$, $R^2 = 0.0464$), indicating that the STG explains 4.64% of the variability in flight time to first solo. The sample was re-analyzed after the identification and removal of outliers. This strengthened the correlation slightly, but overall, it remains a weak negative linear relationship ($R = -0.321$, $R^2 = 0.103$), indicating that the STG explains 10.3% of the variability in flight time to the first solo. While a negative correlation was observed between STG and flight time to first solo, there was not sufficient evidence to demonstrate statistical significance at an alpha level of .05, and as a result, the null hypothesis for RQ2 could not be rejected. Although the results for RQ2 were unable to statistically demonstrate previous findings that PCATD training decreases flight hours prior to the first solo (Vaden et al., 1998), they are consistent with findings from Olson (2002).

The strengths of this research are the methodology accounts for the differences in training experience between domestic and international students through discrete samples, with a subsequent independent statistical analysis of each. This allows the unique external factors influencing each group to be contained, helping to limit cultural effects so as to not skew results, allowing for a clearer view of the relationship between flight simulator performance and training outcomes. The considerable total sample size obtained (n = 195) has allowed for statically significant conclusions to be drawn.

Limitations

Research data were curated from existing data, originally collected for teaching and learning purposes. As such, this prevented control over data collection methods and potential confounding variables. The dataset yielded strong evidence from RQ1 and suggested that international students with better simulator performance were more likely to achieve solo with a fixed-time training syllabus. While this relationship was not statistically demonstrable for domestic students, data did reflect the same general trend of better simulator performance being related to reduced flight training hours to solo flight. The failure of this relationship to be observed with statistical significance in this study may have been due to differences in the flight training circumstances of international and domestic students. One key difference in these training pathways is the time interval between when domestic students undertook simulator training and when their practical flight training began. This data was not gathered or accounted for in this study, which for domestic students is approximately one year. International students, however, commenced their practical flight training course within four to six months after completing the simulator training. It is possible that these delays had an appreciable impact on the strength of findings, as benefits of PCATD training and any skills acquired may have eroded due to the delay, a phenomenon that is well understood in the aviation industry (Martinussen & Hunter, 2017), with inexperienced pilots particularly susceptible (Childs & Spears, 1986). Also, given the comparatively modest TER of PCATD, the skills developed may be less robust than other training methods and may deteriorate against a time delay. This effect may have been further exacerbated through other delays resulting from COVID in 2019, 2020, and 2021 cohorts, along with other traditional contributors to flight training delays, such as weather. In addition, the research only consisted of univariate analysis, and additional factors such as grade-point average (GPA) or participant age were not considered. GPA is a particular factor of interest, as STG may just be a proxy for GPA, and research results are consistent with the previous studies on flight training and GPA (Jones, 2013). Also, whilst a comparatively large sample for this type of study was obtained, it was nonetheless a convenience sample, and this may limit the generalization of research findings to the wider university flight training environment.

Different levels of motivation may also have played a role in performance. International students only had one opportunity to complete their flying course, whether they were able to fly solo or not, and whether students flew solo impacted their ability to continue training independently. A limitation specifically relating to RQ1 is that some students who failed to fly solo may have been limited by the syllabus hour cap and may not have necessarily been a reflection of an inability to achieve solo without further flight hours. Domestic students had greater flexibility in the pace of their training and undertaking remedial flights, which may have affected the pressure on them to perform. These differences in the training circumstances of the two groups may also help explain the overall lower average time to fly solo for international students. Further research can address these factors to unpack the nature of the relationship between simulator performance and practical performance in a flight training context.

Further Research

Key future research avenues of interest concern better understanding the effects of delays between the simulator and commencement of training, reviewing additional prediction methods using simulator performance, better understanding the challenges of flight training as an international student, and multivariate analysis.

Firstly, an investigation into the effect of conducting simulator training closer to the commencement of flight training would help understand the effect of delay between commencement between events. Potentially the more marginal TER of PCATD simulators sees a less sustained training benefit that erodes after a significant delay period, as was the case in this study.

Another potential area of research interest could be to explore other variables that describe simulator performance to further unpack its predictive ability as well as other aspects of training that simulator proficiency can impact. One approach could be to explore the influence of simulator performance on the achievement of various flight training milestones within a certain amount of flight time. Another avenue could be to measure other aspects of simulator performance (such as time to master maneuvers, decision-making capacity, or communication skills) to establish any predictive quality they may offer.

It is noteworthy that only 37% of the international student cohort did successfully complete a solo flight. As to the remaining 63%, further research can explore why the solo flight was unachievable, considering this group still managed to achieve a passing STG. This could in part due to the unique challenges are faced by international students undertaking flight training which can affect outcomes, in particular, the language barrier that is often present (Baugh & Stolzer, 2018). Factors that may not have been relevant to domestic students, such as language or cultural barriers, may provide insight into this disparity.

From a methodological perspective, this study consisted of univariate analysis. Exploration of additional factors such as grade-point average (GPA), participant age, or specific attributes of mental processing such as perceptual speed may provide more insight as to training performance. Also, whilst a comparatively large sample for this type of study was obtained, it was nonetheless a convenience sample, and this may limit the generalization of research findings to the wider university flight training environment.

Conclusion

The use of simulators to support flight training will continue to be of great importance to the aviation community. A better understanding of the complex relationship between simulator training and pilot proficiency can help to ensure appropriate resource allocation and student support.

Results of this study have reinforced the importance of this relationship consistent with findings from the literature (Taylor et al., 1999, as cited in Martinussen & Hunter, 2017, Vaden et al., 1998, Olson, 2002). Results from RQ1 suggest evidence of a relationship existing between an international student's mean STG and whether they achieved the first solo within the fixed syllabus hours. While the results for RQ2 found a negative correlation between STG and flight time to solo, this finding was insufficient to demonstrate statistical significance. As a result, this study suggests a variety of conceptual and methodological avenues for further research to further clarify this relationship.

References

Baugh, B. S., & Stolzer, A. J. (2018). Language-Related Communications Challenges in General Aviation Operations and Pilot Training. *International Journal of Aviation, Aeronautics and Aerospace, 5*(4), 8. https://doi.org/10.15394/ijaaa.2018.1271

Bjerke, E., & Healy, M. (2010). Predicting student persistence: Pre-entry attributes that lead to success in a collegiate flight program. *Collegiate Aviation Review, 28*(1), 25–41. https://doi.org/10.22488/okstate.18.100399

Childs, J. M., & Spears, W. D. (1986). Flight-Skill Decay and Recurrent Training. *Perceptual and Motor Skills, 62*(1), 235–242. https://doi.org/10.2466/pms.1986.62.1.235

Emery, B. (2011). *Neurocognitive predictors of flight performance of successful solo flight students* [Doctoral dissertation, Northcentral University]. ProQuest Dissertations & Theses. https://access.library.unisa.edu.au/login?url=https://search-proquest com.access.library.unisa.edu.au/docview/912202569?accountid=14649

Ennis, E. A. (2009). *The evolution of simulation in aviation flight training* [Master's thesis, California State University]. ProQuest Dissertations Publishing. https://www-proquest-com.access.library.unisa.edu.au/dissertations-theses/evolution-simulation-aviation-flight-training/docview/305181151/se-2?accountid=14649

Goetz, S., Harrison, B., & Voges, J. (2015). The use of FAA flight training and aviation training devices at UAA institutions. *Collegiate Aviation Review, 33*(1), 44-59. http://www.uaa.aero/docs/Spring_2015_CA.pdf

Hoermann, H. J., & Goerke, P. (2014). Assessment of social competence for pilot selection. *The International Journal of Aviation Psychology, 24*(1), 6-28. https://doi.org/10.1080/10508414.2014.860843

Jones, C. A. (2013). *The relationship between academic performance and pilot performance in a collegiate flight training environment* [Master's thesis, Middle Tennessee State University]. ProQuest Dissertations Publishing. https://www-proquest-com.access.library.unisa.edu.au/dissertations-theses/relationship-between-academic-performanceand/docview/1437198712/se-2?accountid=14649

Lutte, R. K., & Mills, R. W. (2019). Collaborating to train the next generation of pilots: Exploring partnerships between higher education and the airline industry. *Industry & Higher Education, 33*(6), 448–458. https://doi.org/10.1177/0950422219876472

Martinussen, M., & Hunter, D.R. (2017). *Aviation psychology and human factors* (2nd ed.). CRC Press. https://doi.org/10.1201/9781315152974

McDermott, J. T. (2005). *A comparison of the effectiveness of a personal computer-based aircraft training device and a flight training device at improving pilot instrument proficiency: A case study in leading regulatory change in aviation education* [Doctoral dissertation, Bowling Green State University]. ProQuest Dissertations Publishing. https://www-proquest-com.access.library.unisa.edu.au/dissertations-theses/comparison-effectiveness-personal-computer-based/docview/305028534/se-2?accountid=14649

McFarland, M. R. (2017). *Student Pilot Aptitude as an Indicator of Success in a Part 141 Collegiate Flight Training Program* [Doctoral dissertation, Kent State University]. OhioLINK Electronic Theses and Dissertations Center. http://rave.ohiolink.edu/etdc/view?acc_num=kent1492088859648498

Mekhail, A., Niemczyk, M., Ulrich, J. W., & Karp, M. (2010). Using the table reading test as an indicator for success in pilot training. *Collegiate Aviation Review*, *28*(1), 101–114. https://doi.org/10.22488/okstate.18.100404

Olson, R. B. (2002). *An analysis of student progress in beginning flight training: Performance prediction, performance measurement, and performance improvement* [Doctoral dissertation, Western Michigan University]. ProQuest Dissertations Publishing. https://www-proquest-com.access.library.unisa.edu.au/dissertations-theses/analysis-student-progress-beginning-flight/docview/276006610/se-2?accountid=14649

Page, R. L. (2000). Brief history of flight simulation. *SimTecT 2000 proceedings* (pp. 11–17). SimTecT 2000 Organising and Technical Committee. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.132.5428&rep=rep1&type=pdf

United States Government Accountability Office. (2014). *Aviation workforce: Current and future availability of airline pilots* (GAO-14-232). https://www.gao.gov/assets/gao-14-232.pdf

Vaden, E. A., Westerlund, K. K., Koonce, J. M., & Lewandowski, W. (1998). The Use of a Personal Computer-Based Aviation Training Device in Ab Initio Flight Training. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *42*(20), 1374–1377. https://doi.org/10.1177/154193129804202002

Zierke, O. (2014). Predictive Validity of Knowledge Tests for Pilot Training Outcome. *Aviation Psychology and Applied Human Factors*, *4*(2), 98–105. https://doi.org/10.1027/2192-0923/a000061

Collegiate Aviation Review International

Volume 40 | Issue 2 Peer Reviewed Article #6

11-6-2022

A Machine Learning Approach Towards Analyzing Impact of Surface Weather on Expect Departure Clearance Times in Aviation

Shlok Misra
Embry-Riddle Aeronautical University

Godfrey Dsouza
Embry-Riddle Aeronautical University

Dothang Truong
Embry-Riddle Aeronautical University

Commercial air travel in the United States has grown significantly in the past decade. While the reasons for air traffic delays can vary, the weather is the largest cause of flight cancellations and delays in the United States. Air Traffic Control centers utilize Traffic Management Initiatives such as Ground Stops and Expect Departure Clearance Times (EDCT) to manage traffic into and out of affected airports. Airline dispatchers and pilots monitor EDCTs to adjust flight blocks and flight schedules to reduce the impact on the airline's operating network. The use of time-series data mining can be used to assess and quantify the impact of surface weather variables on EDCTs. A major hub airport in the United States, Charlotte Douglas International Airport, was chosen for the model development and assessment, and Vector Autoregression and Recurrent Neural Network models were developed. While both models were assessed to have demonstrated acceptable performance for the assessment, the Vector Autoregression outperformed the Recurrent Neural Network model. Weather variables up to six hours before the prediction time period were used to develop the proposed lasso regularized Vector Autoregression equation. Precipitation values were assessed to be the most significant predictors for EDCT values by the Vector Autoregression and Recurrent Neural Network models.

Recommended Citation:
Misra, S., Dsouza, G. & Truong, D. (2022). A machine learning approach towards analyzing impact of surface weather on expect departure clearance times in aviation. *Collegiate Aviation Review International,* 40(2), 79-102. Retrieved from http://ojs.library.okstate.edu/osu/index.php/CARI/article/view/9356/8433

Commercial air travel in the United States (US) has grown significantly in the past decade (2010-2019) (Department of Transportation, 2020). An increase in air traffic in the National Airspace System (NAS) leads to delays and higher operating costs for airlines (Federal Aviation Administration [FAA], 2018). As per the FAA, flight delays are documented under five causes which are carrier delay, late arrival delay, NAS delay, security delay, and weather delay. Weather delays account for the largest cause of flight delays in the US and are the factor for nearly 70% of flight delays in the US (FAA, 2021). Airlines operate with constrained resources and schedule flights based on fixed block times (Sohoni et al., 2017). Delays lead to block time deviations, which can significantly affect the operating network and dispatch operations of an airline. The cost of an hour of flight delay is estimated to be about $1,400 to $4,500 per flight for an airline with the value of passenger time estimated to be in a range of $35 to $63 per hour (FAA, 2021). Airline dispatchers rely on updated air traffic information such as Expect Departure Clearance Times (EDCT) to plan and manage flights and mitigate disruptions to the overall airline network.

A large number of airlines operate in a hub-and-spoke network where the airline's operating network is characterized by single hub or multiple hub airports that are connected to several spokes or connecting airports (Parsa et al., 2019). Airlines develop schedules to ensure passengers traveling within the network can connect to different flights through different hubs (Abdelghany & Abdelghany, 2019). Airlines schedule flights to minimize connection times for passengers and ensure efficiency in the hub airports. For airlines operating in a hub-and-spoke network, EDCT for flights arriving into a hub airport can significantly affect the operations of the entire network due to passenger misconnections, lack of ground equipment, and delays to subsequent flights for the delayed aircraft. EDCTs usually affect flights at specific time banks, which are affected by factors including but not limited to weather, airport capacity constraints, or runway closures (FAA, 2009). Extended EDCTs can lead to extensive delays, ground stops, flight crew limiting on flight duty periods, and flight cancellations. As a possible mitigation tool, delay forecasting is used by airline management to predict the impact of independent factors such as weather events on whether a flight will be delayed (Etani, 2019; Goodman & Griswold, 2019). Airlines invest considerable resources in improving the efficiency of their operational network. An accurate delay forecasting model, such as the model developed in this study, can aid an airline in forecasting EDCTs and planning.

Literature Review

Traffic Management Initiatives (TMI) and Expect Departure Clearance Time (EDCT)

Traffic Management Initiatives (TMIs) are used by Air Traffic Control (ATC) to manage air traffic based on excess demand or a lowered acceptance rate at a particular airport (FAA, 2009). Terminal TMIs are airport-specific initiatives that impact arrivals into a particular airport (FAA, 2009). Some of the common TMIs are Ground Delay Programs (GDPs), Airspace Flow Programs (AFPs), EDCT, and Ground Stop (GS) (FAA, 2009). Non-compliance with a TMI can lead to holding and diverting for aircraft as well as extensions of GDPs, AFPs, and GSs, which leads to further delays due to the overabundance of airplanes, unused slots at destination airports, and increased volume in the airspace (FAA, 2009).

EDCT, a type of TMI, is a runway release time assigned to an aircraft by ATC due to applicable TMIs, which require the aircraft to hold on the ground at their departure airport (FAA, 2009). When EDCTs are assigned to aircraft, the flight crew is given a time window within which the flight is expected to depart (FAA, 2009). EDCTs can be changed based on the conditions at the affected airports, such as changing weather conditions and airport acceptance rate (FAA, 2009). Airline dispatchers monitoring their respective flights provide updated EDCTs directly to company personnel, while pilots can also receive their modified EDCT times from the ATC at the departure airport (FAA, 2009). Like all TMIs, EDCTs are highly influenced by the weather conditions at the airport the flights are scheduled to arrive at (Swot et al., 2018).

Flight Delays Forecasting in Aviation

Flight delay forecasting can be operationalized through different statistical techniques. However, due to the advancements in machine learning algorithms, various studies have focused on forecasting flight delays utilizing machine learning techniques. Machine learning techniques have been demonstrated to be effective for flight delays prediction (Belcastro et al., 2016; Khan et al., 2021; Khanmohammadi et al., 2016; Rebollo & Balakrishnan, 2014; Yu et al., 2019). Khan et al. (2021) utilized a hierarchical integrated machine learning model to predict the flight delays and flight durations for an airline based in Hong Kong. The authors utilized a dataset provided by an airline that consisted of flight data for 19,105 flights and contained data on the runway configuration for the departure airports, weather variables such as atmospheric pressure, air temperature, altitude for flight, speed of the flight, ramp weight of the flight, and type of aircraft. The dataset was regarded as a cross-sectional dataset, and the delays were predicted as a classification problem. The authors developed a Convolutional Neural Network which was named a hyperparameter-free cascade principal component least-squares neural network (hyp-free CPCLS). The hyp-free CPCLS was capable of determining the hyperparameters, such as the number of neurons and layers, without the need for manual hyperparameter tuning. The model was designed due to the highly unbalanced, high dimensional, and highly skewed dataset that was used for the modeling. The authors determined that "categories such as passenger and baggage handling, aircraft and ramp handling, air traffic flow restriction, and government authority, and reactionary and miscellaneous are the main reason for airline departure delay" (p.21). While the study by Khan et al. (2021) contributed in literature to modeling using skewed, high dimensional, and unbalanced datasets through re-sampling and feature engineering techniques, Rebollo & Balakrishnan (2014) focused on capturing the spatial and temporal dependency of departure delays data. While departure delays research focuses highly on local spatial variables for the departure airports, the authors focused on new network delays variables that could impact the entire NAS. The authors defined spatial variables such as NAS Delay State and Type of Delays Day along with temporal variables such as Time of the Day, Month of the Year, and Day of the Week. The authors' work was considered novel due to the focus on including variables that not only impacted the airports for analysis but also impacted the NAS at large. The final model created was a Random Forest model for a 2-hour forecasting period with an average test error of 19%.

Yu et al. (2019) utilized a combination of Deep Belief Networks and Support Vector Repressors to predict flight delays on city-pair routes in China. Yu et al. (2019) explained different feature selection techniques that can be used to develop robust prediction models from high dimensional data. The authors replaced macro-level factors that are commonly seen in flight delay prediction models with specific micro-level influential factors such as aircraft capacity, boarding options, number of passengers in the flight, airline properties, and delay of previous flight for the aircraft. Yu et al. (2019) emphasized the need for feature selection techniques to reduce the dimensionality of datasets by using two conventional filter methods like the Correlation Coefficient Method and the Standard Deviation Selection Method. The final Deep Belief Network-Support Vector Repressor model was able to predict flight delays with a Mean Absolute Error of 8.41, Root Mean Squared Error of 12.65, and Coefficient of Determination of 0.93. The authors determined that air traffic control, delay of the previous flight, and air route situation were the most significant independent variables for the model.

Belcastro et al. (2016) utilized data mining to predict arrival delays due to weather conditions. The authors of the study utilized flight information such as origin airport, destination airport, scheduled departure and arrival times, and weather observations at the departure and arrival airports. The arrival delays prediction was processed as a classification task. The authors developed Decision Tree, Support Vector Machine, Random Forest, Stochastic Gradient Descent, and Naïve Bayes classifiers. The authors evaluated the scalable parallel version of the Random Forest to be the best predictor that could predict arrival delays at a threshold of 60 minutes with an accuracy of 85.6% and a recall of 86.9%. The authors also tested the model with only flight information as predictors and removed the weather conditions predictors, which reduced the model accuracy to 69.1%. In another similar study, Khanmohammadi et al. (2016) examined literature in the field of machine learning models to predict flight delays and examined the role of nominal independent variables in skewing model performance. Khanmohammadi et al. (2016) proposed an Artificial Neural Network that utilized a new type of multi-level input layer to capture the relationship of nominal independent variables. The authors designed a Neural Network model with a multi-level input layer designed for defect of module prediction. The model was deployed to predict flight delays at New York-John F. Kennedy International Airport and was compared to a Gradient Descent Backpropagation model with the same dataset. Some of the nominal independent variables used for the model included the day of the month, day of the week, origin airport, delay at departure at the origin airport, and scheduled departure time. The authors evaluated that model developed was robust to nominal independent variables and was able to predict the flight delays with a Root Mean Squared Error of 0.1366 as compared to 0.1603 for the Gradient Descent Backpropagation model.

Temporal Nature of Flight Delays

As reviewed, machine learning techniques have been successfully utilized for flight delay prediction. However, flight delay data has been modeled differently by scholars. Flight delay data can be treated as cross-sectional data, time-series data, or even spatial data. Determining the data type and format is crucial while deciding the modeling strategy for a machine learning model. While Khan et al. (2021), Belcastro et al. (2016), Khanmohammadi et al. (2016), and Yu et al. (2019) modeled the data as cross-sectional, Rebollo & Balakrishnan (2014) modeled the flight delays utilizing the temporal and spatial dependencies of the variables. Time series forecasting utilizing the temporal dependency of variables has been demonstrated to be an effective method for delay forecasting in aviation (Guvercin et al., 2021; Lan & Shangheng, 2020). Guvercin et al. (2021) used a combination of time series clustering and time series forecasting techniques to build a prediction model to predict flight delays at 305 airports in the US. For the time series forecasting, the authors utilized "a combination of a regression and an Autoregressive Integrated Moving Average (ARIMA) model" (Guvercin et al., 2021, p. 1). As the study was not based on the data for a single airport, the authors needed to utilize a Clustered Airport Model approach to improve forecasting accuracy for the 305 airports. The authors evaluated that the ARIMA approach in combination with the Clustered Airport Model provided forecasting results comparable to forecasting results expected from a complex Long Short Term Memory (LSTM) neural network model. While Guvercin et al. (2021) utilized a clustered airport model to develop a prediction model that could be used for a large number of airports, Lan & Shangheng (2020) collected data from a single "large airport" for four years to develop a model to predict hourly departure delays (p.1). While the hourly departure delays variable contained continuous values, the authors utilized K-means clustering to cluster the delay variable into five categories or bins. For the prediction, the authors determined that Vector Autoregression (VAR) in comparison to Autoregressive Conditional Heteroskedasticity (ARCH) was an effective time series forecasting technique for delay forecasting. While Guvercin et al. (2021) and Lan & Shangheng (2020) were successful in utilizing autoregression models, Zen et al. (2021) utilized a deep graph-embedded LSTM neural network approach for airport delay prediction. A deep graph-embedded LSTM approach was preferable because the authors aimed to develop a model that was based on the data from 325 airports in the US. The authors described the use of the graph-embedded network as a "directed graph network with an airport as a node, a spatial distance weighted adjacency matrix and a demand weighted adjacency matrix are constructed, and the two are integrated to obtain a combined weighted adjacency matrix" (Zeng et al., 2021, p. 13).

Machine Learning Approach for Delay Prediction

The advancement of machine learning techniques has allowed their usage and deployment in tasks across different fields, including aviation. Carvalho et al. (2020) aimed to review the different approaches used by scholars for flight delay predictions from a data science perspective. The authors explored the use of machine learning techniques for flight delay prediction and concluded that the most popular machine learning techniques included k-Nearest Neighbors, Neural Networks, Support Vector Machine, Fuzzy Logics, and Random Forest. The choice of model depends on the prediction, purpose of the project, and data structure. For aviation delay prediction datasets, it is important to preserve the temporal dependencies of variables. Qu et al. (2020) demonstrated the use of Convolutional Neural Networks for time series flight delay prediction. For the modeling process, the authors fused meteorological data and concluded that flight delay prediction accuracy could be improved by up to 1% when using weather data in comparison to predictions by only using flight information. The authors utilized the Airline On-time Performance Database provided by the Bureau of Transportation Statistics in the US for the flight information and Local Climatological Data provided by the National Climate Data Center in the US. While Recurrent Neural Networks are mostly associated with temporal data, Convolutional Neural Networks, as standalone models or in conjunction with any other model, are common for time-series predictions due to their ability to extract the most significant features. The authors utilized a Dual-channel Convolutional Neural Network and Squeeze and Excitation-Densely Connected Convolutional Network for the study. The Dual-channel Convolutional Neural Network and Squeeze and Excitation-Densely Connected Convolutional Network were able to achieve accuracies of 92.1% and 93.19%, respectively.

Research Questions

This study aimed to answer the following research questions:

RQ1: Can EDCT values be predicted for a large hub airport in the US using surface weather observations?

RQ2: What variables are the most significant predictors of EDCT values?

Significance of the Study

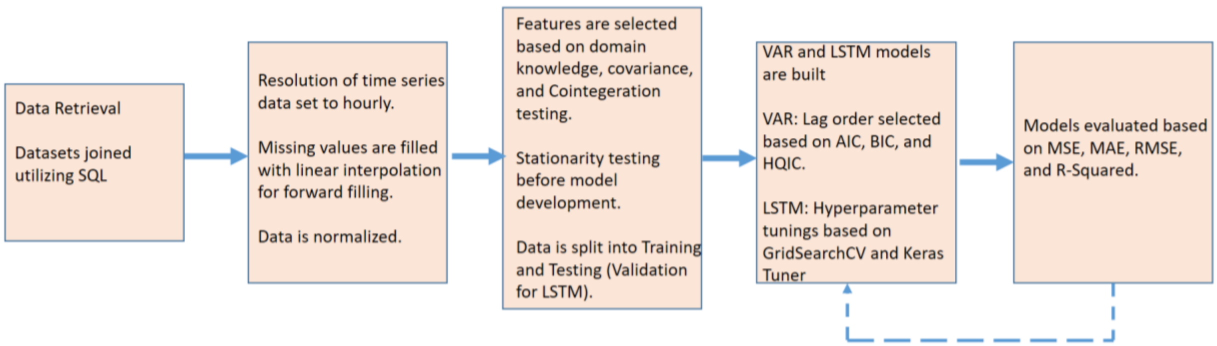
The literature reviewed highlighted the viability and success of machine learning models in predicting different types of flight delays. The effect of surface weather on delays, including EDCTs, has been studied by scholars in the past (Belcastro et al., 2016; Qu et al., 2020). EDCT, just like other TMIs, is severely affected by weather and can disrupt traffic flow for an airport. The studies by Guvercin et al. (2021) and Lan & Shangheng (2020) demonstrated the effectiveness of time series autoregressive models in predicting flight delays. However, there is a significant gap in research in predicting EDCTs utilizing any type of statistical modeling, even though there is domain importance and need for such a prediction model. This study is an attempt to bridge the research gap by utilizing surface weather variables to develop time series models to predict EDCT values for a major hub airport. Time series models will allow the model to retain the temporal dependency of the endogenous variables, which has been demonstrated to be an important concept in published literature.

An EDCT prediction model will allow airline management to make better informed short-term operations decisions such as contingency fuel and resource and gate allocations. EDCT prediction will also help airline management with customer service as longer EDCT predictions can be treated as a direct indication of higher arrival delays at a hub airport which can lead to passengers' missed connections. Additionally, unlike delay parameters such as arrival delays, block delays, and departure delays, EDCTs are issued and enforced by ATC with little to no control by airline management. Based on domain expertise, the scope of EDCT prediction for enhanced airline management and planning is immense, and this study is aimed at adding literature to the subject.

Methodology

The purpose of the study was to develop a time series model to predict EDCTs based on surface weather observations for a large hub airport in the US. Based on the reviewed literature, the researchers adopted Vector Autoregression and Recurrent Neural Network, specifically Long Short Term Memory, modeling approaches for the study. The researchers aimed to develop a VAR model and an LSTM model and compare model performance to predict the EDCTs. For the modeling, the researchers used Charlotte Douglas International Airport (Charlotte). Charlotte is the largest hub for American Airlines in the US, with 397,983 departures and arrivals in 2020 (Charlotte Airport Media, 2021). While the model was built based on the data for Charlotte Douglas International Airport, the researchers expect the results of the study to be transferable for prediction and analysis at other large hub airports as well. Figure 1 depicts the overall model development pipeline adopted for the study.

Figure 1
Overall Model Development Pipeline for the Study



Data Collection and Preprocessing

The researchers acquired historical hourly surface weather observations and hourly traffic data, including EDCT data for Charlotte Douglas International Airport. Two databases were provided by the National Oceanic and Atmospheric Administration (NOAA) and FAA for the weather and traffic information, respectively (FAA, n.d.; NOAA, n.d.). The hourly weather and traffic data for Charlotte Douglas International Airport from 2014-2019 was used in this study. The data for 2020 was included due to the effects of the COVID-19 pandemic on air travel. Once the data was downloaded in comma separate values (CSV) formatted files, the researchers formed a dataset from the different data files using a Structured Query Language (SQL) application with the date/time column as the foreign key. Since the data was structured with data points corresponding to every hour, it could be treated as a time series dataset for the data preprocessing and data analysis stages. The dataset required significant preprocessing due to missing values for some data points. The researchers utilized the Pandas library for the Python programming language for the preprocessing tasks and a forward-filling method to handle missing values. Once the data was preprocessed, it was used to build the VAR and LSTM models.

Vector Autoregression Architecture

Vector Autoregression is a statistical technique used to capture the dependencies of multiple time series variables and the temporal dependencies over time. VARs have been extensively developed and deployed for multivariate time-series predictions. VARs can be used to develop multiple simultaneous equations with the time-lagged values of all the variables, called endogenous variables, used to model and analyze the relationship between the different variables. The VAR model for the study was built utilizing the Statsmodels library in the Python 3.0 Programming Language. Figure 2 depicts the model development strategy developed by the

VAR model. The VAR model can be represented by Equation 1.

$$Yt = \alpha + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \cdots + \beta_p Y_{t-p} + \varepsilon_t \quad (1)$$

$$\alpha = constant$$

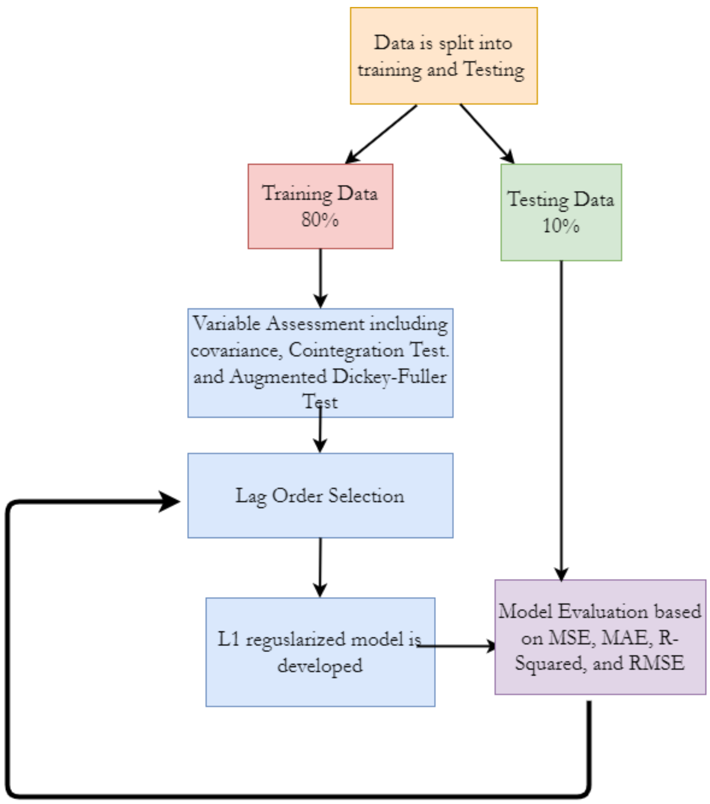
$\beta_1, \beta_2 \ldots \beta_p = Coefficients\ of\ the\ lags\ from\ t\ to\ t-p$

$Y_t, Y_{t-1} \ldots Y_{t-p} = Endogenous\ Variables$

$\varepsilon_t = Error\ Term$

Figure 2

Model Development Strategy for the VAR model



Stationarity Testing for VAR

Autoregressive models perform most effectively when the time series variables exhibit stationarity (Abdulnasser, 2004). The researchers utilized the Augmented Dickey-Fuller Test to test the stationarity of the time series at a significance level of 0.05 (Kulaksizoglu, 2005). The Augmented-Dickey Fuller Test tests the null hypothesis that a unit root is not present in the time series analyzed. Based on the test statistic of the test, which is a negative number, the null hypothesis can be rejected and determined that the unit root is present.

Table 1 depicts the results of the Augmented Dickey-Fuller Test. Based on the results of the test, all the time series variables were determined to be stationary and could be used for the model development without any further adjustments. Figure 3 is a heatmap of the covariance matrix of the variables utilized for the VAR model. The heatmap depicts the covariance between each pair of variables for a given random vector. The covariance matrix can be used to analyze the interrelation of all the individual random variables in the matrix and used in conjunction with the Augmented Dickey-Fuller Test to evaluate any data processing and variable selection needs.

Table 1
Augmented Dickey-Fuller Test

Variable	Test Statistic	Critical Value (0.05)	Number of Lags Chosen	Stationarity
Hourly Arrivals	-20.2629	-2.862	54	Stationary
Hourly Gate Delays	-21.0193	-2.862	55	Stationary
Altimeter	-18.4218	-2.862	55	Stationary
Temperature	-7.7599	-2.862	54	Stationary
Precipitation	-30.8831	-2.862	30	Stationary
Hourly Relative Humidity	-19.8074	-2.862	51	Stationary
Hourly Visibility	-22.411	-2.862	52	Stationary
Average EDCT	-24.7523	-2.862	50	Stationary

Order Selection

Order selection is a crucial aspect of developing a time series model. While tools such as autocorrelation function (ACF) or partial correlation functions (PACF) can be used to determine the appropriate order, the researchers used a combination of the Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), and Hannan-Quinn Information Criterion (HQIC) to determine the appropriate lag order. Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), and Hannan-Quinn Information Criterion (HQIC) are estimators of prediction errors for a statistical model. AIC (Equation 2), BIC (Equation 3), and HQIC (Equation 4) can be used as indicators of the qualities of a model in comparison to other models and can be used for model selection.

$$AIC = 2k - 2\ln\left(\hat{L}\right) \tag{2}$$
$k = number\ of\ estimated\ parameters\ in\ the\ model$
$\hat{L} = Maximum\ value\ of\ the\ likelihood\ of\ the\ function$

$$BIC = k\ln(n) - 2\ln(\hat{L}) \tag{3}$$
$\hat{L} = Maximum\ value\ of\ the\ likelihood\ of\ the\ function$
$n = number\ of\ data\ points$
$k = number\ of\ estimated\ parameters\ in\ the\ model$

$$HQIC = -2L_{max} + 2k\ln(\ln(n)) \tag{4}$$
$L_{max} = Log - Likelihood$
$n = number\ of\ data\ points$
$k = number\ of\ estimated\ parameters\ in\ the\ model$

The researchers used a loop algorithm in the Python Programming Language to determine the AIC, BIC, and HQIC for VAR models with lag orders ranging from 1 to 50. Based on the AIC, BIC, and HQIC evaluation, a lag order of 13 was determined to be the optimal lag order. Once the lag order was determined, the researchers developed the VAR model based on the parameters selected.

Data Preparation

Once the initial statistical testing was completed, the researchers split the data for the training and testing of the model. The total dataset consisted of 52,582 instances or rows, with each row representing an hourly interval. Sci-Kit Learn library on Python was used to split the data with 80% of the data used for training and 20% of the data used for testing. Finally, the researchers set the Shuffle to False to ensure that the temporal order was maintained during the splitting operation. The training data had 47,323 instances, and the testing data had 5,259 instances.

Regression Equation and L1 Regularization

The VAR model developed to predict EDCT would consist of 104 independent variables (for EDCT prediction) due to eight time series variables and a lag order of 13. Such a complex model would increase model cost, complexity, sensitivity to noise or outliers, and the possibility of overfitting (Tan et al., 2019). The researchers utilized the L1 regularization (Lasso) technique to regularize the model and reduce the number of independent variables for the model. Such a regression model is expected to exhibit high performance with lower cost, complexity, and low possibility of overfitting. L1 regularization computation can be illustrated by Equation 5.

$$Cost = \sum_{i=1}^{n}(Y_i - \sum_{j=1}^{p} X_{ij}\,\beta_j)^2 + \lambda \sum_{j=1}^{p} I\beta_j I \tag{5}$$

where $\lambda \sum_{j=1}^{p} I\beta_j I$ is regarded as the penalty term, which is the absolute value of the magnitude of the coefficients.

Recurrent Neural Network

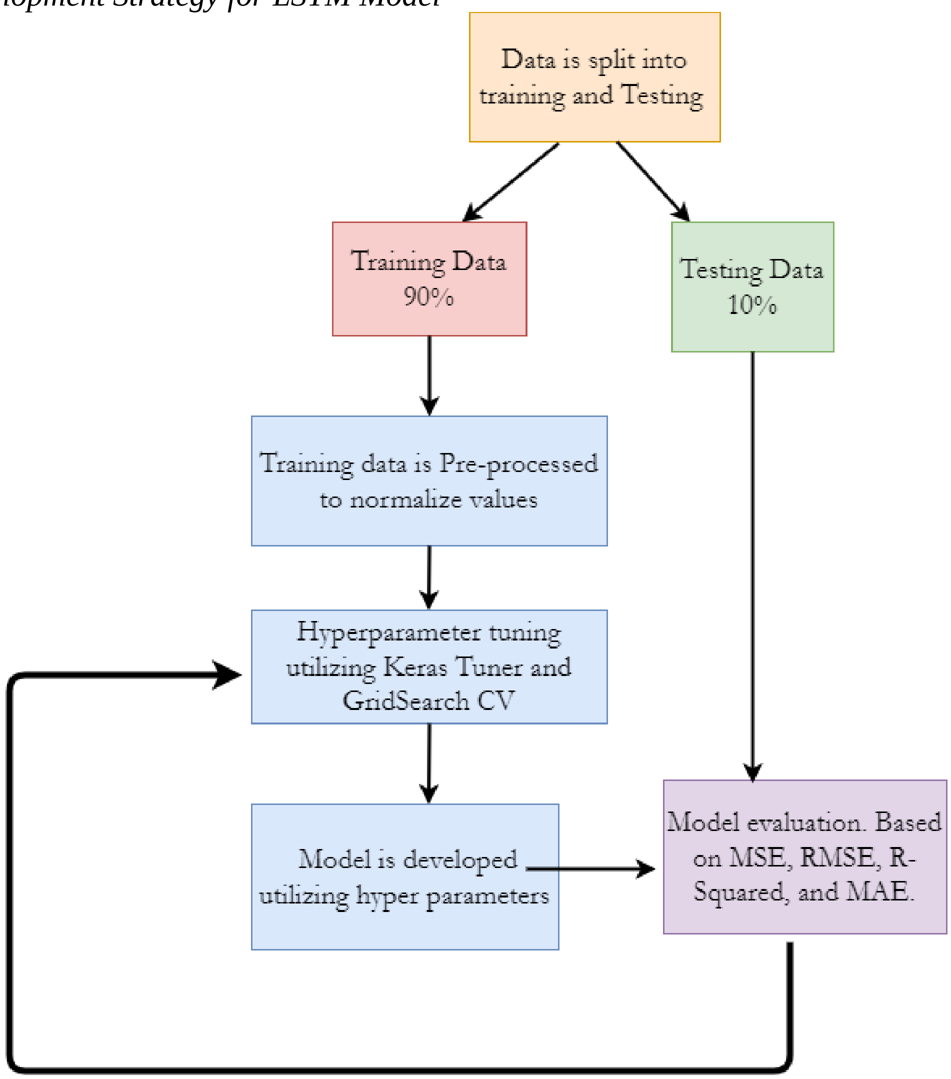
Recurrent Neural Networks are a type of neural network commonly used to model sequential or time series data. Applications of Recurrent Neural Networks include Natural Language Processing, Time Series prediction, Signal Processing, speech recognition, and language translation (Geron, 2019). Long Short Term Memory models are a type of Recurrent Neural Network with the presence of ‘gates’ that are useful for combatting issues such as vanishing and exploding gradients and short-term memory commonly seen in normal Recurrent Neural Networks. With the presence of a Forget Gate, Input Gate, and Output Gate in every LSTM neuron in the network, the model is able to retain long-term memory and dependencies for sequential or temporal data (Geron, 2019). The LSTM model for the study was built using the Tensorflow library in the Python 3.0 Programming Language.

Data Preparation

To build the LSTM model using the TensorFlow library, the researchers needed to preprocess the data into a special array format utilizing the TimeSeriesGenerator library on Python. The researchers utilized the Sci-Kit Learn library to conduct the Train-Test Split operation and set the Shuffle to False to maintain the temporal order of the dataset. For the LSTM model, a validation set was used for the hyperparameter tuning. The data was split with 80% of the data used for training, 10% of the data used for validation, and 10% of the data used for testing. With a total of 52,582 instances, the training dataset had 42,066 instances, the validation dataset had 5258 instances, and the testing dataset had 5,259. The testing dataset for the LSTM and VAR models was the same.

The researchers intended to create a sliding LSTM model and train the model in batches. The window length for the LSTM was set to four, batch size to 32, and sliding to 1. This could be seen as each batch consisting of 32 data points, with each data point containing 4 hours of data with a sliding operation of 1 step. Figure 3 depicts the model development strategy used for the LSTM model.

Figure 3
Model Development Strategy for LSTM Model



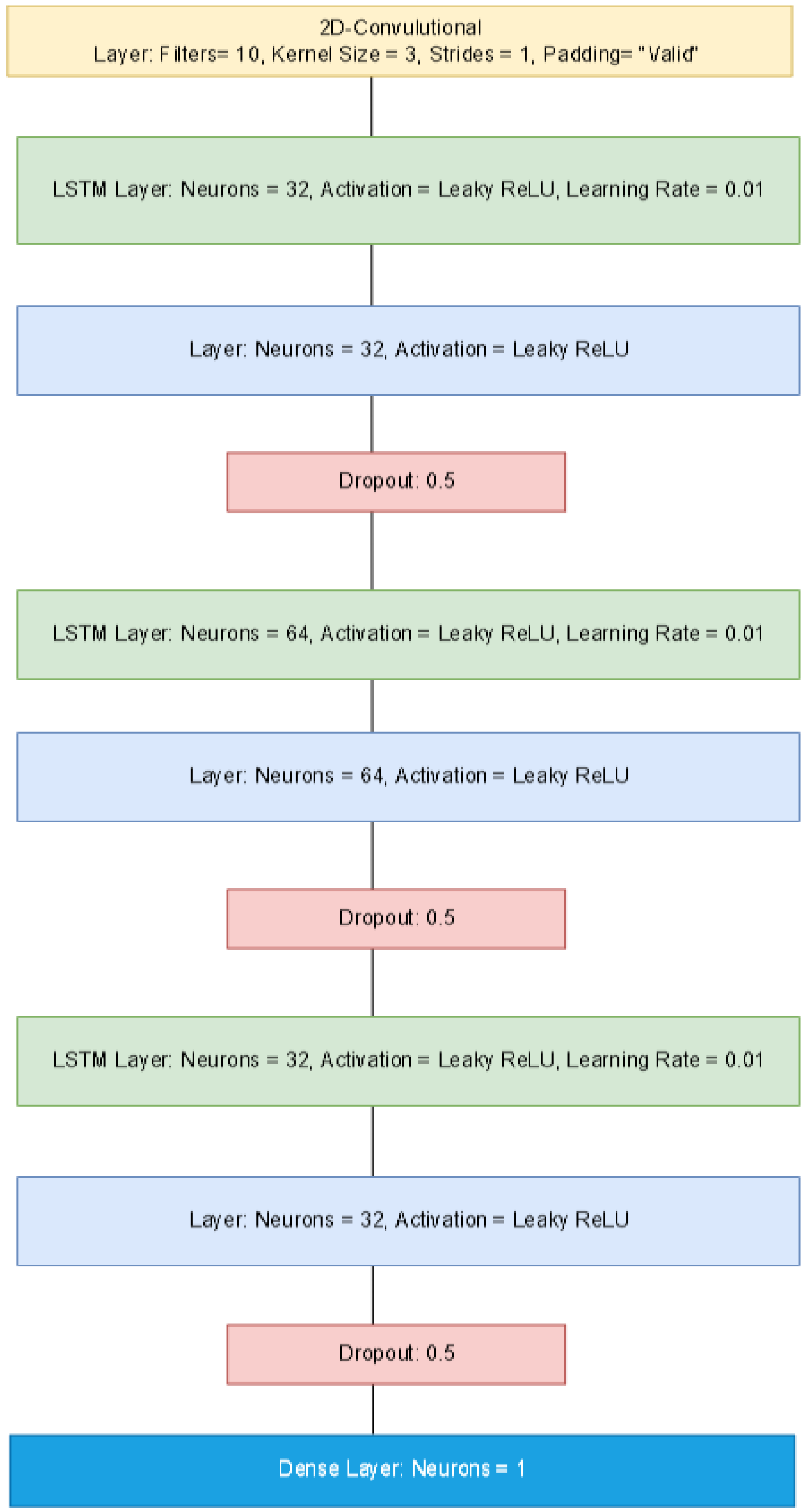
LSTM Architecture and Hyperparameter Tuning

The LSTM model architecture was severely inspired by previous literature on similar prediction tasks. Once the initial model architecture was tuned, the researchers tuned the hyperparameters of the model using the Keras Tuner. A 2D-Convolutional layer was used as the first layer, followed by three LSTM layers with the Leaky Rectified Linear Unit (ReLU) as the activation function. Each LSTM layer was followed by a 50% dropout layer as a regularizer. Additionally, the optimizer was set to Adaptive Momentum (Adam), and the loss function was the mean squared error. Early stopping of the training was added as an additional regularizer. Figure 4 is the model summary output from the Tensorflow library that describes the layer type, activation function, output shape, and parameters for each layer. Figure 5 is an illustration of the LSTM model developed for the study.

Figure 4
LSTM Model Parameters

```
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 conv1d (Conv1D)             (None, 1, 10)             220

 lstm (LSTM)                 (None, 1, 32)             5504

 leaky_re_lu (LeakyReLU)     (None, 1, 32)             0

 dropout (Dropout)           (None, 1, 32)             0

 lstm_1 (LSTM)               (None, 1, 64)             24832

 leaky_re_lu_1 (LeakyReLU)   (None, 1, 64)             0

 dropout_1 (Dropout)         (None, 1, 64)             0

 lstm_2 (LSTM)               (None, 32)                12416

 dropout_2 (Dropout)         (None, 32)                0

 dense (Dense)               (None, 1)                 33

=================================================================
Total params: 43,005
Trainable params: 43,005
Non-trainable params: 0
_________________________________________________________________
```

Figure 5
LSTM Model Parameters



Model Interpretability

The LSTM model was a deep neural network model intended for the regression prediction of the EDCT values. However, for any machine learning model, model interpretability is an important aspect rather than treating the developed model as a *black box*. Shapley Additive Explanations (SHAP) can be used to assess the features utilized to develop a machine learning model, especially a neural network model (Molnar, 2021). Derived from a game theory approach to explain the output of models, SHAP computes Shapley Values utilizing coalitional game theory by treating each feature as a *player* in the game. The SHAP computation can be illustrated by Equation 6.

$$g(\acute{z}) = \emptyset_0 \acute{+} \sum_{j=1}^{M} \emptyset_j \acute{z}_j \tag{6}$$

Where g is the explanation model, $\acute{z} \in \{0,1\}^M$ is the coalition vector, M is the maximum coalition size, and $\emptyset_j$ is the feature attribution of a feature j. A significant advantage of utilizing SHAP to interpret a model is the robustness of SHAP to attribute dependency. As feature importance and permutation importance methods are poor in capturing attribute dependency among the attributes or features used for the model development, they might over-emphasize or under-emphasize some features depending on how those features correlate with other features, which is commonly referred to as the high-correlation variable problem (Hooker et al., 2019). Utilizing Shapely Value Imputation, SHAP is robust to the multicollinearity among the features (Lipovetsky & Conklin, 2001; Lundberg & Lee, 2017). The mean magnitude of SHAP values will be derived utilizing the SHAP library in Python. While the TimeSeriesGenerator library was used to develop the LSTM models on TensorFlow, the training set had to be formatted to a 3D-Array format utilizing the NumPy library due to the limitations of the SHAP library.

Results

Vector Autoregression Model

Based on the data preprocessing, variable selection, stationarity testing, and lag order selection procedures, a VAR model with an order of 13 was developed. Table 2 summarizes the results of the VAR model.

Table 2
Vector Autoregression Model Results

Parameter	Value
Number of Equations	8
Akaike Information Criterion	7.22924
Bayesian Information Criterion	7.57313
Hannan-Quinn Information Criterion	7.33783
Final Prediction Error	1379.17
Log-Likelihood	-627489

The original VAR model built using the Statsmodel library on Python does not involve any sort of regularization. The original VAR model was modified with L1 regularization to remove non-significant endogenous variables for predicting the EDCTs. Finally, a regression equation was developed to predict the EDCT. Table 3 depicts the coefficients, standard error, T-statistic, and probability value associated with each of the endogenous variables used for the regression equation.

Table 3

EDCT Regression Equation Analysis

Variable	Coefficient	Standard Error	T-Statistic	p-value
Lag: 4 Precipitation	25.27544	1.446	17.480	<0.001
Lag: 3 Precipitation	19.0414	1.443	13.194	<0.001
Lag 2: Precipitation	15.2144	1.439	10.566	<0.001
Lag 5: Precipitation	4.1695	1.451	2.873	0.004
Lag 1: Precipitation	2.8705	1.3760	2.086	0.007
Lag 1: EDCT	0.3204	0.0045	70.872	<0.001
Lag 3: Hourly Visibility	0.216	0.0532	4.059	<0.001
Lag 2: EDCT	0.112	0.0047	23.789	<0.001
Lag 1: Hourly Visibility	0.1029	0.0456	2.254	0.004
Lag 6: Temperature	0.0856	0.0265	3.253	0.001
Lag 1: Relative Humidity	0.039	0.0089	4.392	<0.001
Lag 2: Gate Delay	0.006	0.0018	3.295	0.001
Lag 3: EDCT	0.014	0.0046	3.032	0.002
Lag 4: EDCT	0.015	0.0046	3.145	0.002
Lag 5: Precipitation	4.24	1.414	3.003	0.003
Lag 6: Temperature	0.014	0.025	3.147	0.004
Lag 3: Temperature	-0.076	0.0255	-2.679	0.007
Lag 1: Precipitation	3.31	1.342	2.466	0.008
Lag 1: Hourly visibility	0.102	0.044	2.315	0.009
Lag 3: Gate Delays	3.004	1.498	-2.066	0.04
Constant	14.979	10.5503	1.420	0.156

Figure 6
Correlation Matrix of Residuals from the VAR Model

	Hourly Arrivals	Hourly Gate Delays	Altimeter	Temperature	Precipitation	Hourly Relative Humidity	Hourly Visibility	EDCT
Hourly Arrivals	1	0.061448	-0.002337	-0.000043	-0.00274	-0.000108	0.002058	-0.010119
Hourly Gate Delays	0.061448	1	0.01008	0.00863	-0.001987	-0.004312	0.004738	0.11692
Altimeter	-0.002337	0.01008	1	-0.273899	-0.015984	-0.24822	0.074359	-0.006302
Temperature	-0.000043	0.00863	-0.273899	1	-0.129173	-0.456269	0.139224	0.001294
Precipitation	-0.00274	-0.001987	-0.015984	-0.129173	1	0.149912	-0.287004	0.03455
Hourly Relative Humidity	-0.000108	-0.004312	-0.24822	-0.456269	0.149912	1	-0.268077	0.017797
Hourly Visibility	0.002058	0.004738	0.074359	0.139224	-0.287004	-0.268077	1	0.005882
EDCT	-0.010119	0.11692	-0.006302	0.001294	0.03455	0.017797	0.005882	1

Once the VAR model was developed and significant endogenous variables were determined, there was a need to inspect the serial correlation of the residuals to ensure there was a minimal correlation in the residuals and that any patterns in the time series were not left unexplained by the VAR model. Figure 7 depicts the correlation matrix for the model residuals. We can see that there is no endogenous variable that exhibits a high correlation of residuals with EDCTs. We can see a negative correlation between Arrivals and Altimeter with EDCT. The strongest correlation of residuals is exhibited by Hourly Relative Humidity and Temperature. Additionally, the researchers utilized the Durbin Watson Test to check for the serial correlation of the residuals and ensure that the model had sufficiently explained the patterns and variances in the time series dataset used. The value of the Durbin Watson Test can vary between 0 and 4, where a value close to 2.00 implies there is no significant serial correlation (Durbin & Watson, 1971). The Durbin Watson Test is utilized to detect autocorrelation at lag 1 for the prediction errors of an autoregressive model. Table 4 depicts the results of the Durbin Watson Test. The Durbin Watson test results in Table 4 and correlation matrix results in Figure 6 ensured that there was no serial correlation of the residuals and that the model had adequately explained the variance in the data. The Durbin-Watson Statistic used in the Durbin-Watson Test can be represented by Equation 7.

$$d = \frac{\sum_{t=2}^{T}(e_t - e_{t-1})^2}{\sum_{t=1}^{T} e_t^2} \qquad (7)$$

$T = Number\ of\ total\ observations$

$e_t, e_{t-1} = Residual\ of\ the\ autoregression$

Table 4
Durbin Watson Test

Attribute	Durbin Watson Statistic
Hourly Arrivals	1.99
Hourly Gate Delays	1.98
Altimeter	1.99
Temperature	2.01
Precipitation	2.0
Hourly Relative Humidity	1.99
Hourly Visibility	1.98
Average EDCT	2.0

Model Evaluation

The VAR model was evaluated on the testing set on evaluation parameters such as mean squared error, mean absolute error, and root mean squared error. Table 5 illustrates the model results.

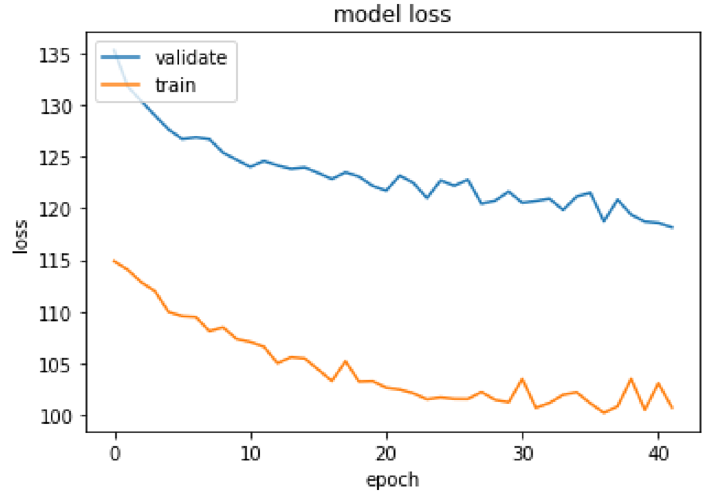
Table 5
Model Evaluation of the VAR model

Evaluation Parameter	VAR model
Mean Squared Error	91.126
Root Mean Squared Error	9.55
Mean Absolute Error	1.99
R-Squared	0.6812

Long Short-Term Memory

An LSTM model was developed on the training test and a validation test. Figure 7 depicts the training and validation model loss with the different epochs. As early stopping was used as a regularizer, the training ceased after epoch 42.

Figure 7
Training and Validation Loss for the LSTM Model with Epochs



The model was evaluated on the evaluation parameters for the training, validation, and testing sets. Table 6 illustrates the model results.

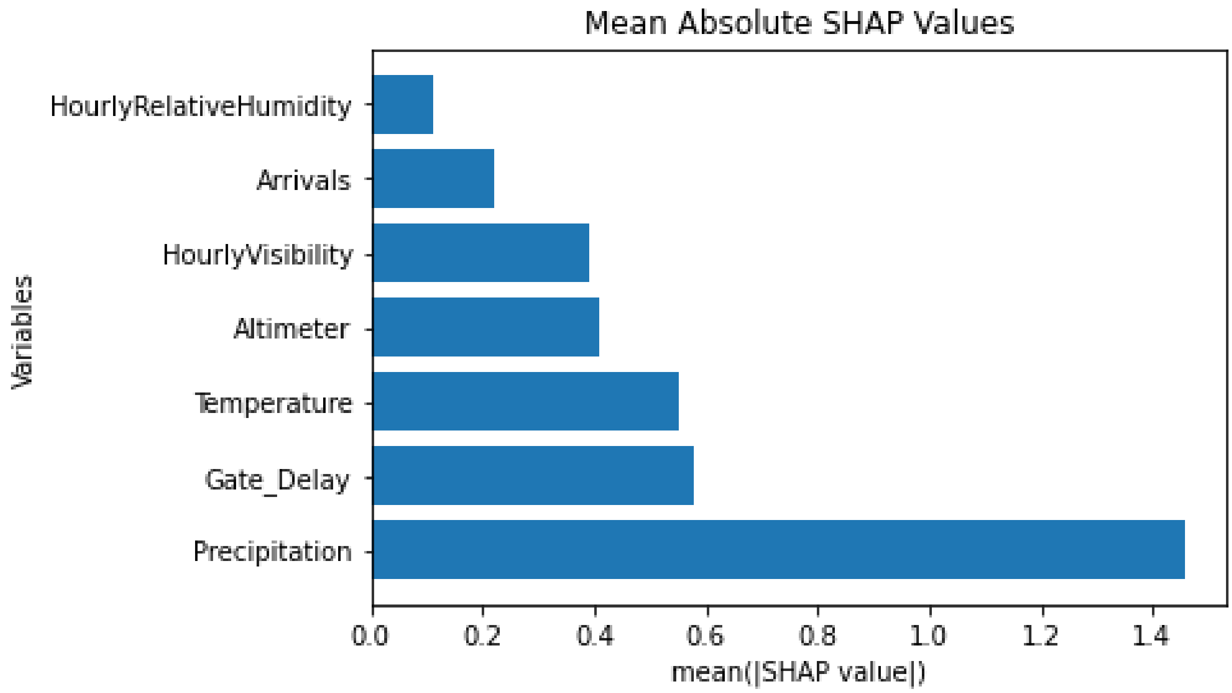
Table 6
LSTM Model Results

Evaluation Parameter	Training	Validation	Testing
Mean Squared Error	102.01	121.11	168.14
Root Mean Squared Error	10.01	11.004	12.96
Mean Absolute Error	2.0073	2.3443	2.85
R-squared	0.681	0.677	0.643

Model Interpretation

SHAP was used to assess the most significant features utilized by the LSTM model for the prediction. The mean absolute SHAP values for the seven features were used to assess their impact on the predictions of the model. Figure 8 depicts the mean absolute SHAP values of the features. Precipitation had the highest mean absolute SHAP value and the highest impact on the target variable followed by Gate Delays.

Figure 8
Mean Absolute SHAP Values



Discussion and Conclusion

The purpose of the study was to develop VAR and LSTM models to predict EDCT for a large hub airport based on surface weather observations. The study was developed based on the significant research gap identified to utilize machine learning techniques to predict EDCTs for an airport, given the importance of EDCTs for dispatch operations of an airline. While there are several demonstrated machine learning algorithms demonstrated by scholars, VAR and LSTM were selected based on previous literature on other domain-related studies. The VAR and LSTM model predictions were primarily evaluated on Mean Absolute Error, Mean Squared Error, Root Mean Squared Error and R-Squared values. While the VAR outperformed the LSTM model on all three evaluation parameters, the performance of both models is considered acceptable. While the LSTM model had lower performance, the researchers believe the results of the LSTM established the viability of utilizing RNNs such as LSTM or Gated Recurrent Units for EDCT predictions. While LSTMs are commonly regarded as more robust time series modeling algorithms due to the non-linear activation and optimization functions involved as compared to VARs, VARs have been demonstrated to outperform LSTMs in previous studies on related subjects (Goel et al., 2016).

While the prediction power of both models was deemed acceptable, it is important to critically analyze the model coefficients or feature importance to understand the most significant predictors. The VAR model can be analyzed based on the endogenous variables coefficients, and the LSTM model can be analyzed based on the SHAP values. The VAR model was regularized with L1 regularization to reduce the number of endogenous variables in the final regression equation and reduce the scope for overfitting the model. It is distinctly clear that precipitation values until Lag 6 had the strongest influence on the EDCT predictions. Additionally, the final regression equation only consisted of endogenous variables up to Lag 6. While the VAR model was built utilizing Lag 13, the L1 regularization reduced the number of endogenous variables because of their low coefficients and, in turn, insignificant impact on the EDCT prediction. The feature assessment for the LSTM was conducted using SHAP. The SHAP analysis is consistent with the coefficients of the VAR model as precipitation was distinctly the highest influencer of EDCT prediction. The model assessments do match intuition as heavy precipitation can be directly associated with convective activity, such as thunderstorms that are a significant cause for delays and in turn issuance of EDCTs.

The model proposed in the study is expected to be a dynamic model in which the input variables are updated hourly for EDCT predictions. Such a model is expected to aid airline dispatchers and airline managers with short-term forecasts and predictions to improve planning and resource allocations. Estimations of EDCT a few hours before the flight can be useful in contingency planning, customer service, and resource allocations at the hub airport. An important utilization for EDCT prediction would be to make necessary adjustments to the airline's contingency fuel policy. In the event of long EDCTs, aircraft return to the gate to obtain more fuel should they go below their minimum fuel required on the dispatch release. Using these predictions, airlines can develop dynamic contingency fuel requirements based on the EDCT estimations. Lastly, the results from the model can help airlines develop and optimize a flight schedule to reduce the number of heavy arrival banks into hubs. Spacing out arrivals among different time banks will reduce the EDCTs and airspace flow constraints.

References

Abdelghany, A., & Abdelghany, K. (2019). *Airline network planning and scheduling*. Wiley Publishing.

Abdulnasser, H. (2004). Multivariate tests for autocorrelation in the stable and unstable VAR models. *Economic Modelling, 21*(4), 661-683. https://doi.org/10.1016/j.econmod.2003.09.005

Belcastro, L., Marozzo, F., Talia, D., & Trunfio, P. (2016). Using scalable data mining for predicting flight delays. *ACM Transactions on Intelligent Systems and Technology, 8*(1), 1-20. https://doi.org/10.1145/2888402

Carvalho, L., Sternberg, A., Gonçalves, L., Cruz, A., Soares, J., & Brandao, D. (2020). On the relevance of data science for flight delay research: a systematic review. *Transport Reviews, 41*(4). https://doi.org/10.1080/01441647.2020.1861123

Charlotte Airport Media. (2021). *CLT welcomes 27.2 million passengers in 2020*. https://cltairport.mediaroom.com/2020-Passenger-Numbers#:~:text=Overall%2C%20aircraft%20operations%20logged%20397%2C983,added%20last%20year%20from%20Charlotte

Cirium. (2015, November 9). *What is "Block Time" in airline schedules? Why does it matter?* https://www.cirium.com/thoughtcloud/block-time-airline-schedules/#:~:text=Block%20time%20includes%20the%20time,t%20break%20these%20elements%20apart

Durbin, J, & Watson, G. (1971). Testing for serial correlation in least squares regression. III. *Biometrika, 58*(1), 1-19. https://doi.org/10.2307/2334313

Etani, N. (2019). Development of a predictive model for on-time arrival flight of airliner by discovering the correlation between flight and weather data. *Journal of Big Data, 6*. https://doi.org/10.1186/s40537-019-0251-y

Federal Aviation Administration. (2009). *Traffic Flow Management in the National Airspace System*. https://www.fly.faa.gov/Products/Training/Traffic_Management_for_Pilots/TFM_in_the_NAS_Booklet_ca10.pdf

Federal Aviation Administration. (2015). *FAQ: Weather delay*. https://www.faa.gov/nextgen/programs/weather/faq/#:~:text=By%20far%2C%20the%20largest%20cause,%22Delay%20by%20Cause%22%20Reports.

Federal Aviation Administration. (2018). NextGen implementation plan 2018-19. *Office of NextGen*. https://www.faa.gov/nextgen/media/ NextGen_Implementation_Plan-2018-19.pdf

Federal Aviation Administration. (2009). *Traffic flow management in the National Airspace System.* https://www.fly.faa.gov/Products/Training/Traffic_Management_for_Pilots/TFM_in_the_NAS_Booklet_ca10.pdf

Federal Aviation Administration. (2021). *Inclement weather*. https://www.faa.gov/newsroom/inclement-weather-0?newsId=23074

Federal Aviation Administration. (n.d.). *FAA operations & performance data.* https://aspm.faa.gov/

Geron, A. (2019). Hands-On machine learning with Sci-kit Learn, Keras, and Tensorflow: Concepts, Tools, and Techniques to build intelligent systems. O'Reilly Publishing. ISBN: 978-1492032649

Goel, H., Melnyk, I., Oza, N., Matthews, B., & Banerjee, A. (2016). *Multivariate aviation time series modeling: VARs vs. LSTMs*. https://goelhardik.github.io/images/Multivariate_Aviation_Time_Series_Modeling_VARs_vs_LSTMs.pdf

Goodman, C., & Griswold, J. (2019). Meteorological impacts on commercial aviation delays and cancellations in the continental United States. *Journal of Applied Meteorology and Climatology, 58*(3), 479–494.

Guvercin, M., Ferhatosmanoglu, N., & Gedik, B. (2021). Forecasting flight delays using clustered models based on airport networks. *IEEE Transactions on Intelligent Transportation Systems*. 22(5). https://doi.org/10.1109/TITS.2020.2990960

Hooker, G., Mentch, L., & Zhou, S. (2019). *Unrestricted permutation forces extrapolation: Variable importance requires at least one more model, or there is no free variable importance*. https://arxiv.org/pdf/1905.03151.pdf

Johansen, S. (1991). Estimation and hypothesis testing of cointegration vectors in Gaussian vector autoregressive models. *Econometrica, 59*(6), 1551-1580. https://doi.org/10.2307/2938278

Khan, W., Ma, H., Chung, S., & Wen, X. (2021). Hierarchical integrated machine learning model for predicting flight departure delays and duration in series. *Transportation Research Part C: Emerging Technologies*, *129*. https://doi.org/10.1016/j.trc.2021.103225

Khanmohammadi, S., Tutun, S. & Kucuk, Y. (2016). A New Multilevel Input Layer Artificial Neural Network for Predicting Flight Delays at JFK Airport. *Procedia Computer Science, 95,* 237-244. https://doi.org/10.1016/j.procs.2016.09.321

Kirchgässner, G., & Wolters, J. (2008). *Introduction to modern time series analysis*. Springer. https://doi.org/10.1007/978-3-540-73291-4

Kulaksizoglu, T. (2015). Lag order and critical values of the augmented dickey-fuller test: A replication. *Journal of Applied Econometrics, 30*(6), 1010-1010. https://doi.org/10.1002/jae.2458

Lan, M, & Shangheng, O. (2020). Characteristic analysis of flight delayed time series. *Journal of Intelligent Systems, 30*(1), 361-375. https://doi.org/10.1515/jisys-2020-0045

Lipovetsky, S., & Conklin, M. (2001). Analysis of regression in game theory approach. *Applied Stochastic Models in Business and Industry, 17*, 319-330. https://doi.org/10.1002/asmb.446

Lundberg, S., & Le,,, S. (2017). A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems 30 (NIPS 2017)*. https://papers.nips.cc/paper/2017/file/8a20a8621978632d76c43dfd28b67767-Paper.pdf

Molnar, C. (2021). *Interpretable machine learning: A guide for making black box models explainable*. https://christophm.github.io/interpretable-ml-book/

National Oceanic and Atmospheric Administration. (n.d.). *Data tools: Local climatological data (LCD).* https://www.ncdc.noaa.gov/cdo-web/datatools/lcd

Parsa, M., Nookabadi, A., Flapper, S., & Atan, Z. (2019). Green hub-and-spoke network design for aviation industry. *Journal of Cleaner Production, 229*, 1377-1396. https://doi.org/10.1016/j.jclepro.2019.04.188

Qu, J., Zhao, T., Ye, M., Li, J., & Liu, C. (2020). Flight delay prediction using deep Convolutional Neural Network based on fusion of meteorological data. *Neural Processing Letters, 52*, 1461-1484. https://doi.org/10.1007/s11063-020-10318-4

Rebollo, J., & Balakrishnan, H. (2014). Characterization and prediction of air traffic delays. *Transportation Research Part C: Emerging Technologies, 44*, 231-241. https://doi.org/10.1016/j.trc.2014.04.007

Sohoni, M., Lee, Y., & Klabjan, D. (2011). Robust airline scheduling under block time uncertainty. *Transportation Science, 45*(4), 451– 464. http://doi.org/10.1287/trsc.1100.0361

Swot, C., Stalnaker, S., & Coats, P. (2018). Simulation-based analysis of early scheduling in the time-based flow management (TBFM) system for flights with expect departure clearance times (EDCT). *AIAA Aviation Forum 2018*. https://doi.org/10.2514/6.2018-3355

Tan, P., Steinbach, S., & Kumar, V. (2019). *Introduction to data mining*. Addison-Wesley Publishing.

U.S Department of Transportation. (2020). Airlines and airports: Traffic. *Bureau of Transportation Statistics.* https://www.transtats.bts.gov /Data_Elements.aspx?Data=2

Yu, B., Guo, Z., Asian, S., Wang, H., & Chen, G. (2019). Flight delay prediction for commercial air transport: A deep learning approach. *Transportation Research Part E: Logistics and Transportation Review*, *125*, 203-221. https://doi.org/10.1016/j.tre.2019.03.013

Zeng, W., Li, J., Quan, Z., & Lu, X. (2021). A deep-graph-embedded LSTM Neural Network approach for airport delay prediction. *Journal of Advanced Transportation, 2021*. https://doi.org/10.1155/2021/6638130

Collegiate Aviation Review International

Volume 40 | Issue 2 Peer Reviewed Article #7

9-7-2022

Enhancing the Aeronautical Decision-Making Knowledge and Skills of General Aviation Pilots to Mitigate the Risk of Bird Strikes: A Quasi-Experimental Study

Flavio Antonio Coimbra Mendonca
Embry-Riddle Aeronautical University

Julius Keller
Purdue University

The purpose of this study was to investigate if a training workshop exploring aeronautical decision-making (ADM) concepts would improve collegiate aviation pilots' knowledge and skills to mitigate the risk of aircraft accidents resulting from bird strikes. Most research and management efforts to mitigate the risk of aircraft accidents resulting from wildlife strikes have focused on airports since empirical data indicate that almost 80% of these strikes occur in this environment. Pilots play an important role in the prevention of wildlife strikes, and research indicates there are opportunities to improve training. Researchers used a one-group pretest-posttest quasi-experimental design. The population of this study consisted of flight instructors and students from two Part 141 four-year degree-awarding collegiate aviation programs. The safety management of wildlife hazards by pilots (N=107) workshop elicited a statistically significant mean increase in the post-test scores ($M = 36.15$, $SD = 5.251$) compared to the pretest scores ($M = 22.29$, $SD = 7.23$), a statistically significant mean increase of 13.858 points, 95% [12.419, 15.298], CI $t(105) = 19.088$, $p < .0005$, $d = 1.85$. The possible benefits of providing Part 141 collegiate pilots with ADM training and education to prevent bird strikes include reducing the direct and other monetary losses resulting from bird strikes and supporting the sustainable growth of the U.S. aviation industry.

Recommended Citation:
Mendonca, F. A. C. & Keller, J. (2022). Enhancing the aeronautical decision-making knowledge and skills of general aviation pilots to mitigate the risk of bird strikes: A quasi-experimental study. *Collegiate Aviation Review International*, 40(2), 103-131. Retrieved from http://ojs.library.okstate.edu/osu/index.php/CARI/article/view/9362/8434

Introduction

Wildlife Hazards to Aviation

On March 04, 2008, a Cessna 500 crashed just after takeoff from Wiley Post Airport (PWA) in Oklahoma City (OK). PWA is a public-use general aviation (GA) airport that is not certificated under Title 14 Code of Federal Regulations (CFR) Part 139. According to the National Transportation Safety Board (NTSB), the airplane was maintaining approximately 200 knots and level at 2,000 feet above ground level (AGL) when it collided with a flock of White Pelicans, one of the largest bird species in North America (NTSB, 2009). Because of the aircraft's airspeed as well as the mass of the bird, "the kinetic energy (KE) resulting from the strikes notably exceeded the airplane certification standards" (Mendonca et al., 2018, p. 5). The aircraft was destroyed by impact forces and a post-crash fire. Both pilots and three passengers were killed because of this accident. Both the Avian Hazard Advisory System (AHAS) (Air Force Safety Center, 2021) and the Federal Aviation Administration (FAA) airport facility directory (FAA, 2022a) entries for PWA contained warning remarks about the high risk of bird strikes at and around PWA. The NTSB (2009), O'Callaghan (n.d.), and Mendonca and Carney (2018) suggested that effective aeronautical decision-making (ADM) by the pilots of this aircraft could have significantly reduced the risk of a mishap resulting from bird strikes.

According to Dolbeer et al. (2022), there were 138,663 reported wildlife strikes to aviation from 2012 through 2021 in the U.S., and almost 5% (N = 6,218) of them caused damage to the aircraft. During the same period, the number of reported wildlife strikes has increased by almost 5%, from 10,643 in 2012 to 15,370 in 2021. Between these years, almost 97% of the total wildlife strikes involved birds. From 1990 to 2021, there were 77 aircraft damaged beyond repair resulting from wildlife strikes. Eighteen of these wildlife strikes resulted in 41 fatalities. Ninety-four percent (N=17) of these strikes involved birds. Moreover, 260 strikes resulted in 337 human injuries. Ninety-three percent (N=238) of these incidents also involved birds. Approximately 97% of the aircraft destroyed were general aviation (GA) aircraft. It is worth mentioning that 56% of those accidents occurred at GA airports.

For commercial and GA aircraft, 82% and 87% of the total wildlife strikes occur in the airport environment (at ≤ 1,500 feet AGL), respectively. Information obtained from the analysis of wildlife strike data indicates that the likelihood of a damaging strike is higher at and above 500 feet AGL. Conversely, considering that almost 97% of the wildlife strikes occur up to 3,500 feet AGL[1], the probability of an aircraft accident resulting from a bird strike is significantly lower when the aircraft is flying above this altitude. While the number of wildlife strikes per one million aircraft movements involving commercial operators increased from 23.92 in 2012 to 33.55 in 2021 (40%), for GA aircraft, this rate increased from 1.69 to 2.80 (66%) during the

[1] For the purpose of this study, the bird-rich zone is defined as the airspace where most bird strikes occur (at ≤ 3,500 feet AGL) (Mendonca et al., 2018).

same period (Dolbeer et al., 2022). The reasons for this growing safety concern at and around GA airports are complex. For example, several GA airports are located in rural areas where hazardous wildlife species to aviation thrive. Moreover, GA airports frequently lack the financial and human resources to develop safety strategies to reduce the risk of wildlife strikes (Cleary & Dickey, 2010).

The FAA requires certificated airports to take safety actions when a triggering event under Title 14 CFR Part 139.337 has been experienced, for example, “an air carrier aircraft experiences multiple wildlife strikes” (Cleary & Dolbeer, 2005, p. 60). Wildlife hazard management plans (WHMP) by airport operators have reduced the number of damaging strikes at certificated airports. Most regulatory and outreach efforts as well as federal funds toward the safety management of wildlife hazards to aviation have been directed Title 14 CFR Part 139 airports. Few federal efforts have targeted the GA community, including airport operators (Cleary & Dickey, 2010). Moreover, GA Airports usually lack financial resources and have limited staff, which limits their capabilities of developing and implementing safety strategies to mitigate the risk of wildlife strikes. Most importantly, analyses of wildlife-strike data have indicated that a multifaceted approach to mitigate the risk of wildlife strikes, especially outside the airport safety management efforts, is vital. Dolbeer et al. (2022) have recommended research, outreach, partnerships, and the use (or innovative use) of new technologies to address this safety hazard afflicting the aviation industry. DeFusco and Unangst (2013), MacKinnon (2004), and Mendonca et al. (2018) have suggested training and education efforts focusing on flight crew members as a strategy to prevent aircraft accidents resulting from strikes.

Aeronautical Decision Making

The FAA (2016a) defines ADM as “a systematic approach to the mental process used by pilots to consistently determine the best course of action in response to a given set of circumstances” (p. 2-1). There are several structured ADM models that can be used by pilots during their ADM processes (.i.e., Pilot, Aircraft, EnVironment, External Pressures [PAVE]). All these ADM frameworks can help pilots identify hazards and mitigate the risks associated with those hazards (FAA, 2009, 2016a). Risk management, an integral component of ADM, is “a decision-making process designed to systematically identify hazards, assess the degree of risk, and determine the best course of action” (FAA, 2008, p. 4). The FAA (n.d.) recommends deliberate training for pilots to gain and/or improve their ADM knowledge and skills. Previous studies (Keller et al., 2015; Li & Harris, 2008; Mendonca et al., 2018) have indicated that ADM is a skill that can be taught. ADM training could be effective in enhancing the risk management, judgment, and decision-making skills of aviation professionals.

Kochan et al. (1997) conducted a series of studies whose goals included the following objectives: to identify and compile ADM skills (i.e., procedural knowledge) that could be used during ADM training of novice GA pilots; and to develop training to improve pilots’ ADM competencies. Findings indicated that ADM expertise could have little association with flight time after a certain number of flight hours (~2,000 hours). The identified ADM skills of “expert pilots” included enhanced situational awareness (Airbus, n.d.) and effective communication. Moreover, findings suggested that hazard awareness and identification as well as risk management knowledge and skills are fundamental for effective ADM processes. Li and Harris

(2001) examined the efficacy of an ADM course in improving the decision-making processes of military pilots in China. Their judgment, risk management, and decision-making processes were evaluated during a series of emergency situations presented in a Northrop F5-E full flight simulator. Their findings suggested that ADM training can improve pilots' risk management knowledge and skills. Keller et al. (2017) investigated the effectiveness of two different training protocols that included ADM and other concepts pertaining to visual flight rules in deteriorating-weather condition encounters. The authors used a pretest-posttest experimental design. The training protocols did not statistically improve the participants' (GA pilots) post-test scores. According to Keller et al. (2017), confounding variables (i.e., previous risk management training) could have biased the result. They also suggested that immersive training could assist in enhancing the pilots' ADM skills to mitigate the risk of weather-related aircraft accidents. O'Hare et al. (2010) conducted a study to examine whether case-based training would improve the ADM process in a simulated flight. The scenarios included encounters with adverse weather. Findings showed that the participants who reflected on a set of cases involving pilots flying into adverse weather conditions and ADM concepts made safer and more timely decisions than other participants who only recalled the material. The authors concluded that case reflection during pilots' ADM training could improve aviation safety. Nonetheless, to the best of our knowledge, few studies have been conducted to examine safety training as a strategy that could help pilots mitigate the risk of aircraft accidents resulting from wildlife strikes. Most importantly, the FAA (n.d.) has encouraged ADM training for pilots incorporating case studies as a strategy to enhance flight safety. According to the FAA (2016a), "the ability to make good decisions is based upon direct or indirect experience and education" (p. 2-3).

Hazards, such as human fatigue, weather, and wildlife, are inherent components of flight operations. Nonetheless, when a pilot is aware of these hazards and follows effective ADM processes, the risks associated with the identified hazards can be reduced and frequently even eliminated. For example, pilots can reduce the probability and the severity of a bird strike by minimizing the flight time and the aircraft airspeed, respectively, while flying through the bird-rich zone (Dolbeer, 2006; Dolbeer et al., 2022; NTSB, 2009). Moreover, previous research studies (Blackwell et al., 2012; Dolbeer & Barnes, 2017; Doppler et al., 2015; Larkin et al., 1975) have suggested that some types of lighting could enhance some bird species' response to an approaching aircraft. Thus, the use of the aircraft's external lights can help reduce the probability of bird strikes involving some bird species (e.g., Canada geese) (MacKinnon, 2004).

Safety Management of Wildlife Hazards to Aviation

According to the FAA (2016a), safety risk management (SRM) is a pillar of ADM. The first step in the SRM process is the identification of hazards (International Civil Aviation Organization [ICAO], 2018). Sources of information about wildlife hazards to aviation include the AHAS (Air Force Safety Center, 2022), the Automatic Terminal Information Service (ATIS), the FAA airport facility directory (FAA, 2022b), Notices to Air Mission (NOTAM) (FAA, 2022c), the aeronautical information manual (AIM) (FAA, 2022d), the airport facility directory (FAA, 2022a), and air traffic control (ATC) (MacKinnon, 2004). It is important to note that the FAA annual and special wildlife-strike reports also provide essential information that must be used to improve aviation safety (Dolbeer et al., 2022). For example, pilots should be aware that the majority of wildlife strikes occur between July and October (migration season) and that the probability of a strike is higher during the arrival phase of the flight (descent, approach, landing roll) (Dolbeer et al., 2022). Nonetheless, the risk of a damaging strike is higher during the departure phases of the flight (takeoff roll and initial climb out). The reasons for that include increasing kinetic energy (KE) (Eschenfelder & Hull, 2006; NTSB, 2009) and a possible loss of thrust at low aircraft energy states and a relatively low altitude in urban areas (Nicholson & Reeds, 2011). Yet, the aircraft is usually heavier (often at maximum takeoff weight) and thus less maneuverable, making evasive actions more difficult. Pilots must understand the reasons why the risk of an aircraft accident resulting from a wildlife strike is higher during the departure phases of flight so that they can exercise their ADM skills and thus improve the safety of their flights.

The aviation system is reliable but complex. Therefore, it is unrealistic to foresee all possible risks involving wildlife hazards. Wildlife strikes, such as occurred with U.S. Airways Flight 1549 (Marra et al., 2009; NTSB, 2010), will require effective ADM processes and immediate actions by pilots using standard operating procedures. Several factors could (e.g., operational constraints by air traffic control) hamper the commercial pilots' ADM processes and resulting a safety risk management process to prevent bird strikes. However, pilots, especially GA aviators, generally have enough time and resources to identify hazards, assess risks, and develop risk management strategies to mitigate risks during flight operations. For example, GA pilots can delay their departure (or arrival) in case of reported bird activity at and/or around the airport. They can also divert to an alternate airport without causing operational and management disruptions in case an air carrier's captain makes a similar decision. "It is important to note that an effective ADM process provides greater latitude for later options, with a significant enhancement of aviation safety" (Mendonca et al., 2018, p. 3).

The safety management of wildlife hazards requires a multifaceted approach. This approach should include airport and aircraft certification standards and guidance materials (Eschenfelder & Hull, 2006), the use of technologies or innovative use of current technologies (Mendonca et al., 2021), and actions by flight crews (MacKinnon, 2004). In fact, pilots play a crucial role in the safety management of wildlife hazards in aviation. For example, pilots can mitigate the severity and/or probability of strikes through appropriate preflight planning, the application of effective ADM processes, and the use of adequate flight procedures (Avrenly & Dempsey, 2013; MacKinnon, 2004). Moreover, pilots have the professional responsibility to report wildlife strikes in accordance with the FAA guidelines (FAA, 2013). Information obtained from the analyses of wildlife strike data provides the scientific basis during the development (and assessment of the effectiveness) of wildlife hazard management plans by airport operators. Additionally, wildlife strike data and information provide the foundation for integrated research as well as FAA national standards and guidance materials to reduce wildlife strikes (Dolbeer et al., 2022). Several initiatives by the FAA with aviation stakeholders (e.g., commercial air carriers) and other government groups have improved the reporting of wildlife strikes under the FAA wildlife-strike report voluntary system as well as the quality of the U.S. national wildlife strike database. However, strike reporting is not consistent across all stakeholders, especially involving the GA community. Thus, aviation stakeholders have advocated for actions that could enhance the quantity and quality (e.g., completeness) of wildlife strikes (FAA, 2013). The various actions undertaken by the FAA, in partnership with other government and aviation industry groups from 2009 to 2013, have enhanced the quantity and quality of reporting of wildlife strikes involving civil aircraft under a voluntary system.

To date, most research and other safety management efforts have focused on airports (Cleary & Dickey, 2010; DeFusco et al., 2015; DeFusco & Unangst, 2013; Mendonca et al., 2020; Rillstone & Dineen, 2013). The safety efforts by airport operators have been successful in mitigating wildlife strikes at the airport jurisdiction. However, such safety strategies have had little effect on wildlife strikes beyond the airport environment (Dolbeer & Barnes, 2017). Additionally, little has been done involving pilots, especially GA aviators. Safety training and education provide aviation professionals with the knowledge, skills, and abilities to carry out their jobs efficiently and safely (ICAO, 2018). Safety training promotes a high level of safety awareness. According to Manuele (2017), safety training should be tailored precisely to the actual needs of those being trained. Additionally, the inherent hazards encountered during flight operations must be considered in the tailoring of the pilot's training. Safety training and education can help pilots develop their ADM competence to identify the inherent hazards of their jobs as well as to develop effective risk mitigation techniques.

Mendonca et al. (2018) conducted a study to investigate if a safety training protocol that included ADM, safety culture, and the safety management of wildlife hazards to aviation concepts would improve Title 14 CFR Part 141 pilots' knowledge and skills to mitigate the risk of bird strikes. The researchers utilized a pretest-and post-test experimental design. Mendonca et al. (2018) also collected qualitative data through open-ended questions in both the pre-and post-test. Findings indicated that the experiment (safety training) significantly increased the post-test scores of the experimental group. A finding of concern obtained from the qualitative data was that the safety management of wildlife hazards by pilots is barely explored during ground and flight training. This study had a number of limitations, including a small sample size (N=17).

The current study built upon Mendonca et al. (2018) and had a similar purpose of investigating if a training workshop exploring ADM applied to the safety management of wildlife hazards would improve the GA pilots' knowledge and skills to mitigate the risk of aircraft accidents resulting from bird strikes. After all, it is practically impossible to mitigate the risk of aircraft accidents resulting from bird strikes without pilots' participation, for example, by effectively applying ADM concepts throughout their flights (Dolbeer, 2006; Eschenfelder & DeFusco, 2010; MacKinnon, 2004; Dolbeer et al., 2022), and/or by reporting wildlife strikes to the FAA (Cleary & Dickey, 2010; FAA, 2013). This quasi-experimental study is in alignment with the FAA's (2022e) overarching objectives of reducing civil aviation and commercial space transportation-related accidents and incidents.

Methodology

Population and Sample

The population of this study consisted of flight instructors and students from two accredited CFR Part 141 flight training and four-year degree-awarding universities, one located in the Midwestern region of the United States and the other one located in Central Florida. This "safety management of wildlife hazards by pilots" workshop was included in the syllabuses of aviation safety-related courses (e.g., crew resource management) delivered during the spring and fall academic semesters of 2020 and during the spring of 2021. All students (N = 107) enrolled in these courses were then required to complete this training as described in the procedures section of this manuscript. It is important to mention that pilots were eligible to participate in this study if they were at least 18 years old, directly involved with their universities' professional flight programs, and if they had flown in the previous six months. Please see Table 1 for the participants' demographics and flight experience information.

Table 1
Summary of Participant Demographic Information

Enrollment Status		
First Year	5	4.6%
Sophomore	13	12.2
Junior	38	35.5
Senior	49	45.8%
Combined Degree Program	2	1.9%
Certifications and Rating Frequencies		
Student Pilot	3	2.8%
Private	18	16.8%
Private - Instrument	5	4.6%
Commercial	23	21.5%
Commercial - Instrument	13	12.3%
Certified Flight Instructor (CFI)	25	23.4%
CFI – Instrument (CFII)	13	12.1%
Multi-Engine	5	4.5%
Multi-Engine Instrument	2	2%
Airline Transport Pilot	0	
Flight Hours		
Minimum	**Maximum**	**Mean**
38	915	226.5

Research Design

Quasi-experimental studies, which usually require fewer resources than true-experimental studies, are appropriate when a control group and/or randomization is not possible (Siedlecki, 2020). “Without either of these, the power of the research to uncover the causal nature of the relationship between the independent and dependent variables is greatly reduced” (Salkind, 2012, p. 230). Nonetheless, quasi-experimental studies are pragmatic because they meet some requirements of causality as well as include participants that are often excluded in true-experimental studies. In the current study, researchers used a one-group pretest-posttest quasi-experimental design (Leedy et al., 2019). The dependent variables were the pre-and post-test scores. The independent variable consisted of a safety training protocol incorporating ADM and the safety management of wildlife hazards concepts. In order to enhance the credibility of the quantitative findings, the researchers added a qualitative section to the post-test (Patton, 2015).

Instruments

The safety management of wildlife hazards by pilots' workshop as well as the questions utilized in both the pre-and post-test, were developed after a thorough literature review of previous studies on the safety management of wildlife hazards in aviation (Avrenly & Dempsey, 2014; Cleary & Dolbeer, 2005; Cleary & Dickey, 2010; DeFusco et al., 2015; Dolbeer, 2020; Dolbeer et al., 2022). Moreover, the review covered ADM concepts (FAA, 2016a), including hazard awareness and identification and risk management. Most importantly, how these concepts are applicable to the safety management of wildlife hazards to aviation by pilots (Eschenfelder, 2005; MacKinnon, 2004; Mendonca et al., 2018; Nicholson & Reed, 2011). Please see the Appendix for the questions utilized in the pre-and post-test.

Cognitive, behavioral, connective, transformative, and constructivist educational learning theories provide instructional designers with a rich source of instructional techniques and strategies that facilitate learning (Western Governors University, 2020). However, no single educational learning theory prescribes a comprehensive set of principles and techniques that could effectively facilitate and influence the learning process in every single context (e.g., aviation; medicine) and/or involving different learners (e.g., adults; seniors). Thus, we incorporated elements of different educational theories during the development and delivery of the workshop. The goal was to create a proper match between the workshop content and objectives as well as the learner's needs and expectations (Ertmer & Newby, 2013).

"The validity of a measurement instrument is the extent to which the instrument measures what it is actually intended to measure" (Leedy & Ormrod, 2005, p. 92). The list of questions utilized in the pre-and post-test underwent a content validation process, as suggested by Sekaran and Bougie (2013). The researchers of this study asked a group of students and faculty members to review the assessment instrument for grammar, syntax, and organization and whether the assessment instrument flowed logically. After addressing the students' and faculty members' suggestions and concerns, we then followed the content validity index (CVI) method to further increase the validity and reliability of the measurement instrument (DeVon et al., 2007; Polit & Beck, 2006). See Mendonca et al. (2018) for further information about the development of the safety management of wildlife hazards by pilots' workshops as well as the processes researchers utilized to validate the measurement instrument.

The pre-and and post-test had every 20 multiple-choice questions (see the Appendix). Each question was worth 2 points. We added a question to the pretest asking participants how often (never / rarely / sometimes / often / always) the safety management of wildlife hazards topic had been addressed during their ground and/or flight training. Additionally, the post-test contained five open-ended questions to help researchers better understand what the quantitative data meant (Patton, 2015). Noteworthy to mention that pretest multiple choice questions were randomly scrambled for the post-test. A Cronbach's alpha analysis for the pretest indicated coefficients of 0.702, an acceptable reliability value. A Cronbach alpha for the post-test indicated a coefficient of 0.878, a high-reliability value.

Procedures

All participation was in accordance with the Institution Review Board (IRB) guidelines. As previously noted, students in aviation safety-related courses (e.g., CRM) during the spring and fall semesters of 2020 and the spring semester of 2021, in which the researchers were also the instructors, were required to participate in the workshop. Participants were asked to complete a quiz (pretest) a week before the workshop. After completion of the pretest, they were reminded that there would be a week dedicated to the topic "safety management of wildlife hazards to aviation," during which there would be a presentation on this and other related topics. Yet, they were expected to review some materials before class to include the Sharing the Skies Manual (chapter 10) (MacKinnon, 2004) and the Pilot Handbook of Aeronautical Knowledge (chapter 2) (FAA, 2016a). They were also told they would be required to complete a quiz after the workshop and that only their highest grade (pretest or post-test) would count towards their final grade in that specific course. The pre-and post-test were completed in a paper format. After transcribing the pretest and post-test grades of the students to an excel spreadsheet using random codes and also posting the students' grades on the learning management system (e.g., Canvas), the researchers returned the students' files (pretest and post-test). Therefore, the participants' results were only identifiable by a randomly assigned code (excel spreadsheet).

The "safety management of wildlife hazards by pilots" had four primary sections, which included an overview of wildlife hazards to aviation, bird-hazard information acquisition and interpretation (Cleary & Dolbeer, 2005; Dolbeer et al., 2021; Eschenfelder & Hull, 2006; Eschenfelder & DeFusco, 2010; Kelly, 2002; MacKinnon, 2004; Mendonca et al., 2018; Nicholson & Reed, 2011), the ADM tenets (especially the safety risk management process) (FAA, 2016a), and a review of information derived from previous studies that could leverage the participants' ADM knowledge and skills to mitigate the risk of aircraft accidents resulting from bird strikes. Elements of existing ADM frameworks, such as the PAVE checklist and the "Perceive, Process, and Perform" (3P) model (FAA, 2016a), were briefly addressed in class. It is important to mention that the ADM models generally help pilots organize their thoughts, become aware of safety hazards, identify hazards before and during flight operations, assess risks, analyze mitigation strategies, and use the most effective safety measures. The hazardous attitudes and respective antidotes regarding wildlife strikes were also debated in class. For example, we emphasized that wildlife strikes could happen to any pilot and that pilots are not helpless since they can take actions (i.e., reduce flight time and/or airspeed while flying through the bird-rich zone if operationally possible) that could help mitigate the risk of bird strikes. Multiple short bird hazard scenarios were presented during the class discussions. The participants were asked to apply their ADM knowledge to mitigate the risk, severity, and/or probability of bird strikes considering these bird-hazard scenarios. For example, we asked, "you are the captain of a single-pilot jet aircraft. You are descending from 10,000 feet to 1,500 feet. Air traffic control (ATC) and other aircraft reported birds at 5,500 feet and below. Which actions could you take to mitigate the risk of an accident due to birds, if operationally possible?" Expected answers were to include reduced flight time and airspeed in the bird-rich zone and the use of the aircraft's external lights. Participants were also expected to report to ATC any other pertinent safety-related information.

In addition, a case study involving a Cessna 500 that crashed after an in-flight collision with large birds was conducted (NTSB, 2009). Participants were asked to, in small groups, identify and explain procedures that pilots could have adopted to mitigate the risk, probability, and/or severity of bird strikes in that specific mishap. During these group discussions, participants were asked to apply the PAVE, 3Ps, and the 5Ps (Plan, Plane, Pilot, Passengers, Programming) ADM models, the risk management process elements (i.e., risk assessment), and the wildlife hazard management concepts (i.e., sources of wildlife hazard information) and explain how that specific accident could have been (hopefully) prevented. The instructors then conducted a brief review and answered the participants' final questions.

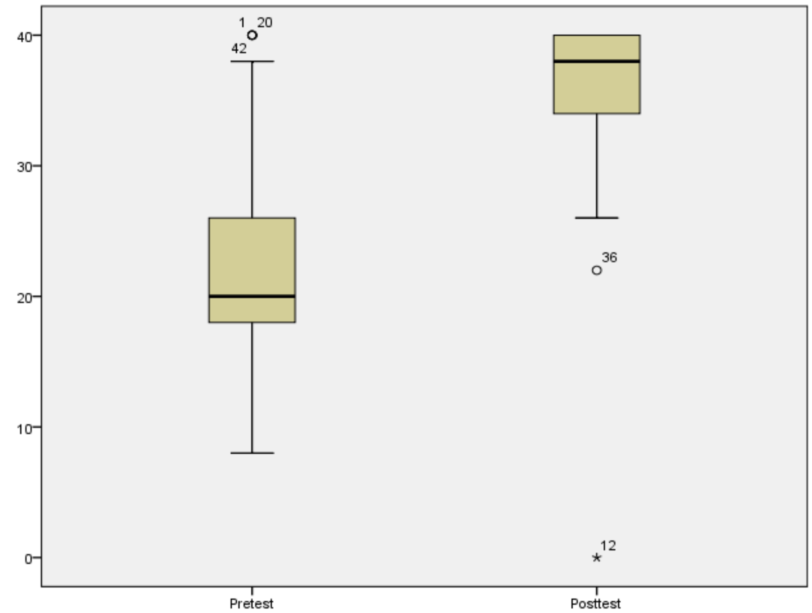
Data Analysis

Descriptive data from the participants were investigated so that researchers could have a better understanding of the data (Salkind, 2012). A paired t-test using SPSS ® was conducted to investigate whether there was a significant difference between the pre-and post-test scores before and after the safety management of wildlife hazards by pilots' workshop. The researchers used the deductive approach to analyze the qualitative data. The goal was to investigate the extent to which the qualitative data support existing ADM and/or safety management of wildlife hazards to aviation concepts (Patton, 2015).

Results and Discussion

Identifying information was not used during the analyses of data, making all research findings anonymous. One hundred and seven students (pilots) from five different courses that occurred in the spring and fall semesters of 2020 and during the spring semester of 2021 participated in the study. Researchers used the paired sample t-test to investigate whether there was a significant mean difference between the participants' pre-and post-test scores. Five outliers were detected that were more than 1.5 box lengths from the edge of the box in a boxplot. However, an inspection of their values did not reveal them to be extreme, and they were kept in the analysis (Laerd, 2021). The difference scores for the post-test and pretest were normally distributed, as assessed by a visual inspection of a Normal Q-Q Plot. The safety management of wildlife hazards by pilots workshop elicited a statistically significant mean increase in the post-test scores (M = 36.15, SD = 5.251) compared to the pretest scores (M = 22.29, SD = 7.23) (see Figure 1), a statistically significant mean increase of 13.858 points, 95% [12.419, 15.298], CI $t(105) = 19.088$, $p < .0005$, d = 1.85.

Figure 1
Comparative Boxplot of the Pretest-Posttest



In the pretest we asked participants, using a five-point Likert rating scale, how often (never / rarely / sometimes / often / always) the safety management of wildlife hazards topic had been addressed during their ground and/or flight training. Most responses indicated they had never (N=12) or rarely (N=51) received any guidance/instruction on that topic during their flight activities. Notwithstanding, 18, 15, and 10 pilots indicated this topic sometimes had, often, or always, respectively, been addressed during ground and/or flight training. The researchers also asked participants if they had taken any academic course addressing the safety management of wildlife hazards to aviation. Only sixteen participants stated they had participated in an academic course that covered this topic. Nonetheless, 37 responses suggested that a focus on pilots' strategies to mitigate the risk of aircraft accidents resulting from bird strikes had not been covered in class. One student stated, "no, besides this course." Another student said, "[…] not in much detail as this class". Another student stated, "I have seen this topic in another course but with a focus on airport wildlife hazard management." A participant said, "[…] first time I had a professor integrate aeronautical-decision making elements and wildlife hazards". Participants were also asked (pretest) if they had had a strike before (and how many). Only16 participants indicated they had experienced bird strikes before, each one of them just a single safety occurrence.

Previous aircraft accidents (FAA, 2022e; NTSB, 2009, 2010) have indicated that wildlife is an inherent risk affecting aviation safety. Nonetheless, effective ADM by pilots (FAA, 2016a) can certainly help mitigate the risks associated with wildlife hazards to acceptable levels (MacKinnon, 2004; NTSB, 2009). According to the FAA (2022d), "it is a pilot's inherent responsibility to be alert at all times for and in anticipation of all circumstances, situations, and conditions affecting the safe operation of the aircraft" (para. 1). Safety training and education to include ADM is a fundamental prerequisite for aviation safety. They ensure aviation professionals develop the skills and knowledge needed to identify, comprehend, and report workplace hazards as well as to develop strategies to effectively mitigate the risks associated with their jobs. Additionally, safety training and education help shape their professional and safety culture. For example, ADM training ensures pilots will have the competencies needed to identify the sources of wildlife hazards to aviation, interpret wildlife hazard data and information, and develop adequate strategies to mitigate the risk of aircraft accidents resulting from strikes. In summary, ADM training and education ensure pilots will have the competencies to safely and effectively perform their duties. Findings of the current study suggest that the topic of safety management of wildlife hazards incorporating ADM concepts had not been frequently addressed during academic courses as well as during flight and/or ground training of the participants of this study.

According to MacKinnon (2004), flight crews can mitigate the risk of wildlife strikes through "prudent flight planning and the use of appropriate aircraft operating techniques" (para. 2). According to Avrenly and Dempsey (2014), pilots should maintain the best angle of climb speed (Vx) in order to reduce both the probability and severity of bird strikes. Dolbeer (2006) suggested pilots should reduce flight time as well as the aircraft airspeed whenever flying in the bird-rich zone in order to reduce the probability and severity, respectively, of bird strikes. In case of a bird encounter, pilots should pull up consistently with good flying techniques since many bird species will dive to avoid an approaching aircraft (Eschenfelder & DeFusco, 2010). Previous research has indicated that the use of the aircraft external lights could enhance the avoidance behavior of certain bird species (Blackwell et al., 2012; Dolbeer & Barnes, 2017).

Qualitative data facilitate the study and understanding of issues in detail and depth. Moreover, they help researchers explain "things that cannot be measured" (Patton, 2015, p. 87). The researchers used the qualitative deductive analysis approach to analyze the qualitative data (Hsieh & Shannon, 2005). This qualitative data analysis methodology can help researchers determine the extent qualitative data explain existing concepts and quantitative results. Yet, demonstrating a link between the quantitative and qualitative researchers can increase the reliability of their study. Participants' responses to the open-ended questions are provided "as is". Thus, they might include grammatical and spelling errors. In the post-test, participants were asked to identify and explain the safest strategy pilots should adopt to mitigate the probability and severity of a bird strike while flying in the bird-rich zone. Almost 80% (N=82) of the participants' responses indicated they would reduce their flight time and airspeed while flying below 3,500 feet AGL to reduce both the probability and severity of wildlife strikes. Some of their answers, which are in alignment with effective ADM concepts to mitigate the risk of bird strikes (Dolbeer, 2015; MacKinnon, 2004), are shown below.

> "Get to the highest altitude in the shortest amount of time."
>
> "Reduce time in the bird-rich zone. Reduce airspeed and increase the angle of descent or climb out at Vx."
>
> "Turn on exterior lights, decrease airspeed, and decrease time in bird rich zone. Turning on the lights lets the birds see you, 'decreases airspeed decreases severity of bird strike, and decrease time in bird rich zone reduces the probability."

> "Pilots should avoid flying in the bird rich zone whenever possible. When it is necessary to do so, climbs should be made at Vx and descents at a high rate of descent without excessive airspeed to minimize time in the bird-rich zone."

> "Less time in bird rich zone, slowdown in bird rich zones to lessen the impact, if you encounter bird pull up on yoke to avoid, use aircraft external lights."

The reporting of wildlife strikes to aviation is voluntary in the U.S. Nonetheless, the FAA (2013) recommends aviation professionals should report strikes following the FAA guidelines. As previously noted, information obtained from the analyses of wildlife-strike data provides the scientific foundation for the safety efforts by aviation stakeholders. Participants were asked why pilots are expected to report wildlife strikes and how they could do it. Ninety-five percent (N=101) of the participants indicated they would report wildlife strikes to the FAA. Interestingly, most of their responses also suggested they generally understood the importance of reporting strikes to the FAA for aviation safety. Some of their answers are as follows:

> "Reporting a bird strike can be good for multiple reasons […]. Please report any bird strikes to the FAA bird strike page on their website".

> "You should report a bird strike to the FAA wildlife strike database and to the ATC controller controlling you".

> "The reporting of bird strikes by pilots enhances safety management […). Reports should be done at the FAA wildlife strike database".

> "Reporting bird strikes is crucial for flight safety. You can report a bird strike through FAA's website".

> "So that more data can be collected by the FAA and future strikes be prevented. Strikes can be reported to the FAA, ATC, and the safety office".

According to the FAA (2016a), pilots should integrate risk management into planning at all levels, including preflight planning. When flight crews follow effective ADM practices, hazards are identified and the associated risks reduced or even eliminated. The first step in the safety risk management process is the identification of hazards. Pilots must, before the beginning of each flight become familiar with all available information concerning that flight (FAA, 2022d). During the workshop researchers explained how pilots can obtain useful wildlife hazard information from those "wildlife hazard" sources, and how this information should be utilized during the pilots' ADM processes. In the post-test, we asked participants what aeronautical resources pilots could use, during their ADM processes, to gather information about the presence of birds at and around airports. All participants indicated they were familiar with key resources they could use to obtain wildlife-hazard data and information during their ADM processes. Some of their responses are shown below.

> "ATC, NOTAMS, ATIS, AHAS, AIM, and FAA wildlife page"

> “They can use the US Air Force Avian Hazard Advisory System, ATC, NOTAMS”.
>
> “The U.S. Avian Hazard Advisory System website is a resource pilots can use to gather information about the presence of birds at their destination or departure airport”.
>
> “Pilots can use resources such as NOTAMS, ATIS, ATC, AHAS, FAA Wildlife strike website, AIM and more”.

Empirical data (Avrenly & Dempsey, 2014; Eschenfelder, 2005) and previous aircraft accidents (NTSB, 2009, 2010) have indicated that the risk of an aircraft accident resulting from a bird strike is higher during the departure phase of flight (takeoff roll and initial climb-out). The faster rotation of the engine as well as the increasing airspeed (both high KE) during these phases of flight can help explain this difference. Researchers explored the KE concept during the safety workshop explaining how pilots should take that into consideration during their ADM processes (e.g., reduce the aircraft airspeed and/or engine rotation while flying through the bird rich zone, if operationally possible, to reduce KE in case of a strike). The KE concept was further explored during the case study involving the Cessna 500 accident (NTSB, 2009).

We asked participants in which phase(s) of flight the risk of an aircraft accident due to bird ingestion(s) is the highest. We also asked them to explain the reasons for that. Eighty-two percent (N=87) of the participants made similar statements. Their responses suggested they understood how the KE concept is important for accident prevention regarding bird hazards. Some of their responses are as follows:

> “The risk of an aircraft accident due to bird ingestions is during takeoff roll and climb out. This is because the engine is at a high-power setting and more likely to cause damage if a bird strike is encountered”.
>
> “During takeoff and climb due to the heavy requirement of thrust needed to complete the phase of flight. Also during these phases of flight the aircraft is near the ground thus leaving less time to glide to safety”.
>
> “Climb is the most dangerous because the aircraft is increasing airspeed and using full power”.
>
> “The highest risk due to bird strikes is during takeoff and initial climb out. This is because airplane engines are running at a high-power setting and can be damaged or totally knocked out while running at this high-power setting. Another reason is that typically the airplane is at lower altitudes and may not have enough gliding distance in the event of a strike”.
>
> “The takeoff/departure and landing/approach are the phases that bird ingestion is highest. This is because this area and altitude is where most birds and wildlife are concentrated in”.

According to the FAA (2016b), air traffic control should issue advisory information for at least 15 minutes on pilot-reported and/or observed bird activity at and around airports to increase the flight crews' awareness of bird hazards. Flight crews should utilize this information to exercise their ADM processes and thus improve aviation safety. In the post-test, we also asked participants why they were expected to inform ATC about the presence of birds they have observed during flight operations. All participants made similar statements. See below some of their answers:

> "Informing ATC about the presence of birds is crucial because ATC then can alert other pilots that might not be aware of birds in the area".
>
> "So that ATC can vector airplanes away from the area".
>
> "So that they can inform other pilots and aircraft in and around the airport and local area of the hazard".
>
> "To enable other pilots to be situationally aware and make safe decisions you should inform ATC about the presence of birds observed while flying".
>
> "This enables ATC to warn other aircraft of the presence of wildlife which can help in mitigating the risk of bird strikes".

In the last open-ended question in the post-test, we asked participants what wildlife mitigation techniques and guidance they had been provided during their careers as pilots and by whom. A few of their responses are as follows:

> "If you see birds when on approach, just continue the approach, they usually dive away – […] my private pilot instructor".
>
> "Birds typically avoid our planes because of our lights, so just follow our normal light SOP and you should have no problem with birds – my instrument Instructor".
>
> "I was taught during my private pilot training outside of XXXX University to climb if you see birds and they will dive; however, other than that I have not been informed about anything related to wildlife mitigation techniques besides this class".
>
> "I have never really been given any guidance to avoiding wildlife mitigation other than not to hit them or to swerve out of their way".
>
> "Speaking truthfully, outside of this XXXX course I had never really been taught much about wildlife mitigation techniques. When presented with birds in the airplane I was simply told to avoid them any way possible".
>
> "I was taught to climb if I encountered birds because most birds tend to dive when they are startled by an aircraft. I have found this advice to be true, and it has helped me avoid multiple bird strikes".

> "While in flight if a bird is approaching you just remain in the same state of flight, don't move, just stay straight and level (flight instructor)".
>
> "My instructors have only ever told me to avoid birds by flying away from the direction they were going".
>
> "My instructor emphasized to pull up if we approach a bird and to use scanning to techniques to prevent a bird strike from occurring".
>
> "I have always been told when approaching birds to keep flying straight because they will move to avoid you, and that if you move you may turn directly into them".
>
> "My instructors have taught me to always be aware of my surroundings especially during arrival and departure. I have been taught to always use all the available resources to be situationally aware of any wildlife in the vicinity".
>
> "The only wildlife mitigation techniques and guidance I have been provided during my flight training were during my aviation safety class, by the professor".
>
> "I was told by my private instructor (not at XXXX University) to always pull up when near birds, because they have a tendency to dive".

Participants' responses to the open-ended questions suggest that their ADM knowledge and skills mitigate the risk of aircraft accidents resulting from strikes improved after the safety training protocol. For example, their responses indicated they understood the importance of reducing flight time and airspeed, if operationally possible, to reduce both the probability and severity of strikes. All participants demonstrated a certain awareness of the key resources pilots could use to obtain bird hazard information that is needed during their ADM processes. Quoting one pilot "make use of the AHAS, AIM, and ATC information to improve your situational awareness and risk management process". All respondents demonstrated a sound understanding of the importance of reporting wildlife hazards to ATC for accident prevention purposes. In summary, their responses to the open-ended questions suggested that after the safety training participants had the ability to think analytically and clearly about the safety management of bird hazards. In addition, their responses to the open-ended questions in the post-test suggested they had adequate knowledge to discuss several mitigation strategies applicable to the safety management of bird hazards by pilots.

Discussion

Training and education are key components of any business plan (Rodrigues & Cusick, 2012), and "safety training is no exception" (p. 319). Effective ADM is a vital aspect of risk management in aviation, without which superior results cannot be achieved. A large amount of research has indicated that ADM is trainable (Buch, 1984; Jensen, 1987; Keller, 2015; Li & Harris, 2001; Mendonca et al., 2018; O'Hare et al., 2009). Pilots frequently make critical decisions to mitigate the inherent risks affecting aviation safety. The analyses of wildlife-strike data in the U.S. have suggested that birds pose a growing economic and safety concern for the aviation industry (Dolbeer et al., 2022). Eschenfelder and DeFusco (2010), Eschenfelder and Hull (2006), Dolbeer (2006), MacKinnon (2004), and Mendonca et al. (2018) have advocated for pilots' training and education as a strategy to mitigate the risk of aircraft accidents resulting from wildlife strikes. Therefore, ADM training focusing of the safety management of birds by pilots is essential to provide those professionals with the required knowledge and skills to make swift and accurate decisions to improve safety.

One hundred and seven pilots from two accredited CFR Part 141 flight training and four-year degree-awarding universities participated in the safety training workshop exploring ADM applied to the safety management of wildlife hazards. Results using parametric tests indicated that there was a statistically significant difference between the pre-and post-test scores of the participants. These findings suggested that training and education focusing on hazard awareness, the safety management of wildlife hazards to aviation, and ADM concepts, as suggested by the FAA (2016a), Keller et al. (2017), and Mendonca et al. (2018) can help prevent aircraft accidents resulting from bird strikes. Additionally, ADM training could improve the quantity and quality of wildlife-strike reports, as suggested by Dolbeer (2015). As previously noted, wildlife-strike data and information is vital for accident prevention efforts by aviation stakeholders.

Participants' responses to the open-ended questions in the post-test generally reflected the "safety management of wildlife hazards to aviation applying ADM" concepts covered during the workshop. For example, when asked about the safest strategy to mitigate the risk of an aircraft accident while flying through the bird-rich zone, their responses echoed the ADM concepts by Dolbeer (2006), Eschenfelder and DeFusco (2010), the FAA (2016a), and MacKinnon (2004). The researchers acknowledge that some of the participants' responses to these open-ended questions, despite adequate (or sometimes incorrect), were not in alignment with was covered during the workshop. For instance, when asked why pilots should report bird strikes, one participant stated, "[…] the airport can start taking immediate action to scare them away. Maintenance can be ready to meet you on the ground". Moreover, when asked about the resources aviation professional could use to obtain wildlife hazard information, one participant incorrectly indicated they could gather such information by consulting a meteorological terminal air report (METAR). Nonetheless, the qualitative findings of this study corroborate the quantitative findings and leverage the idea that a training workshop exploring ADM applied to the safety management of wildlife hazards can certainly improve the safety of the aviation industry.

Two primary themes became apparent during the analysis of the qualitative data. The topic safety management of bird hazards to aviation by pilots incorporating the ADM tenets has not been adequately covered during flight training and education in at least two Part 141 college aviation universities. One participant indicated that this was the first time they had seen a professor explain ADM models focusing on the safety management of bird strikes in aviation. Another theme that emerged from participants' responses to the open-ended questions was that an immersive workshop focusing on the safety management of bird hazards incorporating safety risk management processes and a case study, as suggested by the FAA (n.d.), Kochan et al. (1997), and Mendonca et al. (2018) has the potential to improve aviation safety. More specifically, the safety of the GA community. The participants' responses suggested that they became familiar with the ADM tenets applicable to bird hazard mitigation after the workshop. For example, the participants' responses suggested that they were familiar with key aeronautical resources (i.e., AHAS; NOTAMs) pilots should use to obtain information about the presence of birds at and around airports during their ADM processes. Yet, they understood how the KE and bird-rich zone concepts, as suggested by Dolbeer et al. (2006) and MacKinnon (2004) are important during their ADM processes. Empirical information indicates that high-quality training and education have a greater impact on aviation safety and efficiency than "just the total flight hours accumulated by entry-level pilots" (Mendonca et al., 2021).

A finding of concern, which shed some light on Mendonca et al. (2018) was that the participants of this study had been scarcely provided with guidance to mitigate the risk of bird strikes during flight activities. Yet, much of the guidance provided is not underpinned by empirical information related to ADM concepts applied to the safety management of wildlife hazards in aviation. Training and education are needed for a pilot to gain and/or improve their ADM knowledge and skills. The FAA (2007; 2016a) suggests that case-based and/or scenario-based training can help pilots improve their ADM competencies. We recommend flight instructors should incorporate "bird-hazard scenarios" during ground and flight training in order to help pilots in Part 141 collegiate aviation environment enhance their ADM knowledge and skills to mitigate the risks associated with birds during flight operations. Similarly, we recommend this subject be addressed in academic classes in which ADM concepts are taught to flight students pursuing an academic degree in Part 141 collegiate aviation environment. These recommendations are in alignment with previous studies by Keller (2015), Keller et al. (2020), and Li and Harris (2001).

Conclusion

The aviation industry is a vital contributor to the increasing U.S. economic productivity and prosperity (FAA, 2020a). The U.S. civil aviation system accounts for approximately 5.5% of the U.S. gross domestic product, generates almost $2 trillion in economic activities, and supports more than four million direct and seven million indirect jobs with $490 billion in earnings. The outbreak of the COVID-19 pandemic caused devastating financial and economic losses and significant uncertainties to the aviation industry (ICAO, 2021). Notwithstanding, over the long term, the capabilities and strengths developed by the aviation industry during the last decade will again become evident. Moreover, prospering U.S. and world economies provide the foundation for the U.S. aviation industry to thrive in the next decades. The U.S. commercial aviation industry is forecast to grow by 4.9% per year until 2041. The long-term outlook for the general aviation (GA) industry is also promising. The number of GA hours flown is forecast to increase an average of 0.7 per year until 2040 while the GA fleet is expected to grow by approximately 0.1% during the same period (FAA, 2020b).

Annually, it is estimated that wildlife strikes cost the U.S. civil aviation industry 139,469 hours of aircraft downtime and $328 million in direct and indirect costs (Dolbeer et al., 2022). Empirical information indicates that pilots play a major role in the aircraft accident prevention process. By applying ADM concepts before and during their flights, crews could significantly reduce the risk of a mishap resulting from a strike. The ever-increasing risk of aircraft accidents resulting from wildlife strikes and the optimistic forecast for the U.S. aviation industry require a multifaceted approach to address this safety hazard. This approach should include integrated research, new technologies and/or innovative use of current technologies, and pilots' training and education. The possible benefits of providing Part 141 collegiate pilots with ADM training and education to prevent bird strikes include:

1. Reducing the direct and other monetary losses resulting from bird strikes;
2. Reducing the number of human injuries and fatalities resulting from bird strikes;
3. Increasing the quantity and improving the quality of wildlife strike reports by pilots;
4. Supporting the sustainable growth of the U.S. aviation industry; and
5. Providing these professional pilots unique opportunities to develop or enhance competencies (e.g., risk management) that are valued by the aviation industry.

Limitations and Future Studies

There are a number of limitations associated with this study. The non-probability sampling technique as well as the absence of a control group limits the generalizability of our findings (Sekaran & Bougie, 2013). A convenient sample, for example, is not representative of the entire collegiate flight student population enrolled in accredited CFR Part 141 flight training and four-year degree-awarding universities. The participants of the current study, however, are not unlike their peers in other similar universities. Thus, our findings could provide an interesting picture of the importance of ADM training for accident prevention regarding bird hazards. Another possible caveat to the study findings was the reliability and validity of the assessment tools. Researchers used the assessment instruments utilized by Mendonca et al. (2018). A Cronbach alpha for the pretest and for the post-test of the current study indicated a coefficient of 0.702 (acceptable) and of 0.878 (high –reliability), respectively. Moreover, we supported the quantitative findings with qualitative data as suggested by Patton (2015) in order to increase the credibility of our findings. Nonetheless, we acknowledge that the findings of this study need to be interpreted with caution.

Further research is warranted to better understand why collegiate aviation pilots have received little guidance and instruction on the safety management of wildlife hazards to aviation. Moreover, future studies could investigate if improvement in knowledge regarding the safety management of bird hazards by pilots would be associated with improvements in pilots' practices. We cannot rule out the possibility that some of our findings were biased by demand effects and/or due to other confounding and/or extraneous variables not measured. At last, despite the limitations of this study, findings could provide the scientific foundation for collegiate aviation program leaders' efforts to enhance aviation safety.

References

Airbus (n.d.). *Flight operations briefing notes, human performance: Enhancing situational awareness*. https://www.smartcockpit.com/docs/Enhancing_Situation_Awareness.pdf

Air Force Safety Center (2022). *United States avian hazard advisory system*. https://www.safety.af.mil/Divisions/Aviation-Safety-Division/BASH/

Avrenli, K. A., & Dempsey, B. J. (2014). Statistical analysis of aircraft-bird strikes resulting in engine failure. *Journal of the Transportation Research Board, 2449*, 14-23. https://doi.org/10.3141/2449-02

Blackwell, B. F., DeVault, T. L., Seamans, T. W., Lima, S. L., Baumhardt, P., & Juricic, E. F. (2012). Exploiting avian vision with aircraft lighting to reduce bird strikes. *Journal of Applied Ecology*, *49*, 758-766. https://doi.org/10.1111/j.1365-2664.2012.02165.x

Buch, G., & Diehl, A. (1984). An investigations of the effectiveness of pilot judgment training. *The Journal of the Human Factors and Ergonomics Society*, *26*(5), 557-564. https://doi.org/10.1177/001872088402600507

Cleary, E. C., & Dickey, A. (2010). *Guidebook for addressing aircraft/wildlife hazards at general aviation airports* (ACRP Report No. 32). Transportation Research Board on the National Academies. http://www.trb.org/Publications /Blurbs/163690.aspx

Cleary, E. C., & Dolbeer, R. A. (2005). *Wildlife hazard management at airports: A manual for airport personnel*. Federal Aviation Administration. http://www.faa.gov/airports/airport_safety/wildlife/resources/media/2005_faa_manual_complete.pdf

DeFusco, R. P., & Unangst, E. T. (2013). *Airport wildlife population management: A synthesis of airport practice* (ACRP Synthesis 39). The National Academy of Sciences, Engineering, and Medicine. https://www.nap.edu/catalog/22599/airport-wildlife-population-management

DeFusco, R. P., Unangst, E. T. J., Cooley, T. R., & Landry, J. M. (2015). *Applying an SMS approach to wildlife hazard management* (ACRP Report No. 145). The National Academy of Sciences, Engineering, and Medicine. https://www.nap.edu/catalog/22091/applying-an-sms-approach-to-wildlife-hazard-management

DeVon, H. A., Block, M. E., Wright, P. M., Ernst, D. M., Hayden, S. J., Lazzara, D. J., Savoy, S. M., & Polston, E. K. (2007). A psychometric toolbox for testing validity and reliability. *Journal of Nursing Scholarship, 39*(2), 155-164. https://doi.org/10.1111/j.1547-5069.2007.00161.x

Dolbeer, R. A. (2006). Height distributions of birds as recorded by collisions with civil aircraft. *Journal of Wildlife Management, 70*(5), 1345-1350.

https://digitalcommons.unl.edu/cgi/viewcontent.cgi?article=1496&context=icwdm_usdanwrc

Dolbeer, R. A. (2015). *Trends in reporting of wildlife strikes with civil aircraft and in identification of species struck under a primarily voluntary reporting system, 1990-2013*. Federal Aviation Administration. https://digitalcommons.unl.edu/cgi/viewcontent.cgi?article=1190&context=zoonoticspub

Dolbeer, R. A. & Barnes, W. J. (2017). Positive bias in bird strikes to engines on left side of aircraft. *Human-Wildlife Interactions, 11*(1), 33-40. https://doi.org/10.26077/hp2d-c437

Dolbeer, R. A. (2020). Population increases of large birds in North America pose challenges for aviation safety. *Human-Wildlife Interactions, 14*(3), 345-357. https://doi.org/10.26077/53f9-edc3

Dolbeer, R. A., Begier, M. J., Miller, P. R., Weller, J. R., & Anderson, A. M., & (2022). *Wildlife strikes to civil aircraft in the United States: 1990–2021* (Serial Report Number 28). Federal Aviation Administration. https://www.faa.gov/sites/faa.gov/files/2022-07/Wildlife-Strike-Report-1990-2021.pdf

Doppler, M. S., Blackwell, B. F., DeVault, T. L., Juricic, E. F. (2015). Cowbird responses to aircraft with light tuned to their eyes: Implications for bird-aircraft collisions. *The Condor Ornithological Applications, 117*(2), 165-177. Retrieved from http://www.bioone.org/doi/full/10.1650/CONDOR-14-157.1

Ertmer, P. A., & Newby, T. J. (2013). Behaviorism, cognitivism, constructivism: Comparing critical features from an instructional design perspective. *Performance Improvement Quarterly, 26*(2), 43-71. https://northweststate.edu/wp-content/uploads/files/21143_ftp.pdf

Eschenfelder, P. (2005, May). *High speed flight at low altitude: Hazard to commercial aviation?* (Paper presentation). Seventh Bird Strike Committee USA/Canada Annual Meeting, Vancouver, Canada.

Eschenfelder, P., & Hull, S. (2006, August 21-24). *Reduction of risk: a flight crew guide to the avoidance and mitigation of wildlife strikes to aircraft* (Paper presentation). Eighth Bird Strike Committee USA/Canada Meeting, St Louis, Missouri, United States.

Eschenfelder, P., & DeFusco, R. (2010, August). Bird strike mitigation beyond the airport. *AeroSafety World, 5*(7). https://flightsafety.org/asw-article/bird-strike-mitigation-beyond-the-airport/

Federal Aviation Administration (FAA). (n.d.). *The art of aeronautical decision-making: Course table of contents*. https://www.faasafety.gov/files/gslac/courses/content/28/216/The%20Art%20of%20Aeronautical%20Decision.pdf

Federal Aviation Administration (FAA). (2008). *Aeronautical decision making* (FAA-P-8740-69 AFSH-8083-2). https://www.faasafety.gov/files/gslac/library/documents/2011/Aug/56413/FAA%20P-8740-69%20Aeronautical%20Decision%20Making%20%5Bhi-res%5D%20branded.pdf

Federal Aviation Administration (FAA). (2009). *Risk management handbook* (FAA-H-8083-2). https://www.faa.gov/regulations_policies/handbooks_manuals/aviation/media/faa-h-8083-2.pdf

Federal Aviation Administration (FAA). (2013). *Reporting wildlife aircraft strikes* (Advisory Circular 150/5200-32B). https://www.faa.gov/documentLibrary/media/Advisory_Circular/AC_150_5200-32B.pdf

Federal Aviation Administration (FAA). (2016a). *Pilot's handbook of aeronautical knowledge* (FAA-H-8083-25B). https://www.faa.gov/regulations_policies/handbooks_manuals/aviation/phak/media/pilot_handbook.pdf

Federal Aviation Administration (FAA). (2016b). *Air traffic organization policy: Air traffic control* (JO 7110.65W CHG). https://www.faa.gov/documentlibrary/media/order/atc.pdf

Federal Aviation Administration (FAA). (2020a). *The economic impact of civil aviation on the U.S. economy*. https://www.faa.gov/about/plans_reports/media/2020_nov_economic_impact_report.pdf

Federal Aviation Administration (FAA). (2020b). *FAA aerospace forecast: Fiscal years 2020-2040. https://www.faa.gov/data_research/aviation/aerospace_forecasts/media/FY2020-40_FAA_Aerospace_Forecast.pdf*

Federal Aviation Administration (FAA). (2022a). *FAA airport/facility directory*. https://www.faa.gov/air_traffic/flight_info/aeronav/digital_products/dafd/search/advanced/

Federal Aviation Administration (FAA). (2022b). *Automatic terminal information service procedures.* https://www.faa.gov/air_traffic/publications/atpubs/atc_html/chap2_section_9.html

Federal Aviation Administration (FAA). (2022c). *NOTAMs, TFRs, and aircraft safety alerts.* https://www.faa.gov/pilots/safety/notams_tfr/

Federal Aviation Administration (FAA). (2022d). *Aeronautical information manual. https://www.faa.gov/air_traffic/publications/atpubs/aim_html/*

Federal Aviation Administration (FAA). (2022e). *AVS FY 2022 business plan.* https://www.faa.gov/about/plansreports/avs-fy-2022-business-plan

Federal Aviation Administration (FAA). (2022f). *Some significant wildlife strikes to civil aviation aircraft in the United States, January 1990 – March 2022.* https://www.faa.gov/sites/faa.gov/files/airports/airport_safety/wildlife/significant-wildife-strikes-1990-mar-2022.pdf

International Civil Aviation Organization (ICAO). (2018). *Safety management manual* (Doc. 9859). https://skybrary.aero/sites/default/files/bookshelf/5863.pdf

International Civil Aviation Organization (ICAO). (2021). *Economic impacts of COVID-19 on civil aviation*. https://www.icao.int/sustainability/Pages/Economic-Impacts-of-COVID-19.aspx

Jensen, R. S., Adrion, J., & Lawton, R. S. (1987). *Aeronautical decision making for instrument pilots*. US Department of Transportation. http://www.cfijapan.com/study/ADM_IFR.pdf

Keller, J. (2015). *Unexpected transition from VFR to IMC: An examination of training protocols to mitigate pilot gaps in knowledge and performance* (Publication No. 10099231) [Doctoral dissertation, Purdue University]. ProQuest Dissertations Publishing.

Keller, J., Carney, T., Xie, A., Major, W., & Price, M. (2017). VFR-into-IMC: An Analysis of Two Training Protocols on Weather-Related Posttest Scores. *Journal of Aviation Technology and Engineering*, *7*(1), 2-18. https://doi.org/10.7771/2159-6670.1150

Kelly, T. A. (2002, September). Managing bird strike risk with the avian hazard advisory system. *United States Air Force Flying Magazine: Safety, 58*(9). http://www.usahas.com/Downloads/Article%20Managing%20Bird%20Strike%20Risk%20With%20AHAS%20Flying%20Safety%20Sep%202002%20Kelly.pdf

Kochan, J. A., Jensen, R. S., Chubb, G. P., & Hunter, D. R. (1997). *A new approach to aeronautical decision-making: The expertise method*. https://www.faa.gov/data_research/research/med_humanfacs/oamtechreports/1990s/media/AM97-06.pdf

Larkin, R. P., Torre-Bueno, J. R., Griffin, D. R., Walcott, C. (1975). Reactions of migrating birds to light and aircraft. *Proceeding of the National Academy Science, 72*(6), 1994-1996. *https://www.pnas.org/content/pnas/72/6/1994.full.pdf*

Leedy, P. D., & Ormrod, J. L. (2020). *Practical research: Planning and design*. Pearson Education Limited.

Li, W-C., & Harris, D. (2008). The evaluation of the effect of a short aeronautical decision-making training program for military pilots. *The International Journal of Aviation Psychology, 18*(2), 135-152. https://doi.org/10.1080/10508410801926715

MacKinnon, B. (2004). *Sharing the skies manual – An aviation industry guide to the management of wildlife hazards.* Government of Canada, Transport Canada. http://www.tc.gc.ca/eng/civilaviation/publications/tp13549-menu-2163.htm

Manuele, F. A. (2017). *On the practice of safety*. John Wiley & Sons, Inc.

Marra, P. P., Dove, C. J., Dolbeer, R. A., Dahlan, N. F., Heacker, M., Whaton, J. F., Diggs, N. E., France, C., & Henkes, G. A. (2009). Migratory Canada geese cause crash of US airways flight 1549. *Frontiers in Ecology and the Environment, 7*(6), 297-301. https://doi.org/10.1890/090066

Mendonca, F. A. C., & Carney, T. Q. (2018, March 5-8). *General aviation pilots' strategies to mitigate bird strikes* [Conference session]. *104th Purdue Road School Transportation Conference & Expo*, West Lafayette, IN, United States. https://docs.lib.purdue.edu/roadschool/2018/presentations/49/

Mendonca, F. A. C., Carney, T. Q., & Fanjoy, R. O. (2018). Enhancing the safety training of GA pilots to reduce the risk of bird strikes: An experimental pilot study. *International Journal of Aviation, Aeronautics, and Aerospace, 5*(4), 1-27. https://doi.org/10.15394/ijaaa.2018.1281

Mendonca, F. A. C., Keller, J., & Dillman, B. G. (2021). Competency-based education: A framework for a more efficient and safer aviation industry. *Journal of the International Society of Air Safety Investigators, 54*(1), 19-23. https://www.isasi.org/Documents/ForumMagazines/Forum-2021-JanToMarch.pdf

National Transportation Safety Board (NTSB). (2009). *Crash of Cessna 500, N113SH following an in-flight collision with large birds - Oklahoma City, Oklahoma* (NTSB/AAR-09/05-PB2009-910405). www.ntsb.gov/investigations/AccidentReports/ Reports/AAR0905.pdf

National Transportation Safety Board (NTSB). (2010). *Loss of thrust in both engines, US airways flight 1549 and Subsequent Ditching on the Hudson River: US Airways Flight 1549 Airbus A320-214, N106US Airbus Industry A320-214, N106US* (NTSB/AAR-10/03). http://www.ntsb.gov/investigations/AccidentReports/ Reports/AAR1003.pdf

Nicholson, R., & Reed, W. S. (2011, July). *Strategies for prevention of bird-strike events*. BOEING. https://www.boeing.com/commercial/aeromagazine/articles/2011_q3/4/

O'Callaghan, J. (n.d.). *Bird-strike certification standards and damage mitigation*. National Transportation Safety Board. https://www.ntsb.gov/news/events/documents/oklahoma_city_ok-2_web_bird_strike_cert_and_damage_john_ocallaghan.pdf

O'Hare, D., Mullen, N., & Arnold, A. (2009). Enhancing aeronautical decision making through case-based reflection. *The International Journal of Aviation Psychology, 20*(1), 173-178. https://doi.org/10.1080/10508410903415963

Patton, M. Q. (2015). *Qualitative research & evaluation methods*. SAGE Publications, Inc.

Polit, D. F. & Beck, C. T. (2006). The Content Validity Index: Are you sure you know what's being reported? Critique and recommendations. *Research in Nursing & Health, 29*(5), 489-497. https://doi.org/10.1002/nur.20147

Rillstone, D. J., & Dineen, C. M. (2013). *Airport responsibility for wildlife management* (ACRP Legal Research Digest No. 20). The National Academy of Sciences, Engineering, and Medicine. https://www.nap.edu/catalog/22517/airport-responsibility-for-wildlife-management

Salkind, N.J. (2012). *Exploring research*. Upper Saddle River.

Siedlecki, S. L. (2020). Quasi-experimental research designs. *Clinical Nurse Specialist, 26*(3). 131-135. https://pubmed.ncbi.nlm.nih.gov/32796378/

Sekaran, U., & Bougie, R. (2013). *Research methods for business*. SAGE Publications, Inc.

Western Governors University (2020). *Five educational learning theories*. https://www.wgu.edu/blog/five-educational-learning-theories2005.html#close

Appendix: Questions Utilized during the Assessments

1. Which factor should pilots consider while planning their flights to mitigate the risk of a bird strike?
   a. Birds tend to be more active during the day and dusk
   b. **Birds tend to be more active during dusk and dawn**
   c. Birds tend to be more active during dusk and night
   d. Birds tend to be more active during dawn and day

2. The likelihood of a damaging bird strike is higher:
   a. Below 200 feet AGL
   b. **Above 500 feet AGL**
   c. Above 5,500 feet AGL
   d. Below 500 feet AGL

3. During initial climb out, pilots should clear bird-hazard risk areas by using airspeeds and flap configurations that provide:
   a. The best rate of climb (Vy)
   b. **The best angle of climb (Vx)**
   c. The takeoff safety speed (V2)
   d. The aircraft design maneuvering speed (Va)
      3.1. Why would you use that configuration?

4. The degree of severity resulting from a bird strike is influenced by several factors, especially:
   a. The altitude of the aircraft and its airspeed
   b. The outside air temperature (OAT) and the aircraft configuration
   c. The aircraft airspeed and the mass of the bird
   d. The mass of the bird and the altitude of the aircraft
   e. The phase of flight and the altitude of the aircraft

5. The probability of a bird strike, while flying in a bird-rich zone, is an inverse function of the aircraft:
   a. Airspeed
   b. **Rate of climb**
   c. Configuration
   d. Angle of attack

6. There is some empirical evidence that the following aircraft systems may assist pilots in mitigating the risk of bird strikes involving certain species of birds:
   a. Aircraft external lights and transponder
   b. **Aircraft weather radar and external lights**
   c. Aircraft traffic collision avoiding system (TCAS) and external lights
   d. Aircraft ground proximity warning systems (GPWS) and weather radar

7. The risk of damage to the aircraft engine(s) due to bird ingestions is higher:
   a. **During takeoff run and initial climb-out**

b. During the approach and landing
c. During landing and takeoff run
d. During cruise and descent

8. The bird-rich zone is defined as the airspace where most bird strikes occur. Which actions could aviators take to mitigate the risk of bird strikes in regards to the bird-rich zone?
a. **Reduce flight time and airspeed while flying in the bird-rich zone**
b. Increase flight time and the airspeed while flying in the bird rich zone
c. Increase flight time but decrease airspeed while flying in the bird rich zone
d. Decrease flight time but increase airspeed while flying in the bird rich zone

9. You are cleared by air traffic control (ATC) to descend from 6,000 feet to 1,500 feet and you are aware that you will fly through an area where other pilots reported the presence of birds. Which actions could you take to mitigate the risk of a bird strike, if operationally possible?
a. Increase your rate of descent and the aircraft airspeed
b. **Increase your rate of descent without increasing the aircraft airspeed**
c. Reduce your rate of descent and increase the aircraft airspeed
d. Increase the aircraft airspeed without increasing your rate of descent

10. The severity of a bird strike is a direct function of:
a. The aircraft rate of climb and mass of the bird
b. The aircraft angle of attack and airspeed
c. **The aircraft airspeed and mass of the bird**
d. The aircraft airspeed and rate of climb

11. You are the captain of a single-pilot jet aircraft. You are descending from 10,000 feet to 1,500 feet. Air traffic control (ATC) and other aircraft reported birds at 5,500 feet and below. Which actions could you take to mitigate the risk of an accident due to birds, if operationally possible?
a. Increase airspeed and rate of descent
b. **Decrease airspeed and use idle-power setting**
c. Increase airspeed and use idle-power setting
d. Decrease rate of descent and increase power

12. You are planning a cross-country flight that will most likely pass along a major U.S. migratory flyway. Where can you obtain information about the migration patterns of birds as well as a forecast of bird movements within a low-level flight arena for the contiguous 48 States?
a. The FAA Wildlife Hazard Management Website
b. The U.S. Department of Agriculture (USDA) – Wildlife Services Website
c. The U.S. Bird Strike Committee Website
d. **The U.S. Avian Hazard Advisory System Website**

13. You are flying solo at 2,000 feet when you see a flock of birds at around the same altitude. The recommended action you can take to mitigate bird strikes, if operationally possible, will be:
a. **Pull up, consistent with good flying techniques to attempt to pass over them**
b. Dive, consistent with good flying techniques to attempt to pass below them

c. Turn right or left, consistent with good flying techniques to attempt to avoid them
d. A combination of “B” and “C”

14. You are the captain of a Boeing 787. You are on final approach close to the runway at Chicago O’Hare International Airport when you observe a flock of birds. You realize that your aircraft will probably hit some birds. Which actions would you take to mitigate the risk of bird strikes, if operationally possible?
a. Initiate a go-around and then attempt a second approach
b. **Fly through the birds and land**
c. Initiate a go-around and divert to another airport
d. Fly through the birds and then initiate a go-around

15. You are planning a local flight for the next weekend. You will take some friends on a one-hour ride, and will fly no farther than 35 miles away from KLAF. What can you do to reduce the probability of bird strikes?
a. Fly above 1,000 feet as much as possible
b. Fly below 3,500 feet as much as possible
c. **Fly above 4,500 feet as much as possible**
d. Fly below 2,500 feet as much as possible

16. You are about to takeoff at an airport with reported bird activity. The most effective way for you to reduce the probability of a bird strike after takeoff will be to:
a. Fly at the lowest airspeed possible
b. Minimize the rate of climb
c. Fly at the highest airspeed possible
d. **Maximize the rate of climb**

17. You have just taken off and are about to fly through a bird-rich zone. You should ________ the ________ to reduce the severity of a possible bird strike.
a. Increase / airspeed
b. **Decrease / airspeed**
c. Increase / rate of climb
d. Decrease / rate of climb

18. You will most likely fly through an area where the presence of birds is expected. The following factors could directly impact the amount of damage to the aircraft in case of a bird strike:
a. Engine thrust and aircraft rate of descent
b. Aircraft rate of descent and airspeed
c. **Engine thrust and aircraft airspeed**
d. Aircraft airspeed and configuration

19. Bird strikes during _________ are substantially more hazardous with respect to engine failure than those during _______ and __________.
1. Climb
2. Landing roll

3. Approach
4. Takeoff run
5. Descent
    a. 1, 3, and 5
    b. 5, 1, and 2
    c. 3, 1, and 4
    d. 5, 3 and 4

20. In case you have a bird strike, you should report this incident to the:
    a. **FAA**
    b. National Transportation Safety Board
    c. Chief-pilot
    d. Bird Strike Committee USA
    e. U.S. Department of Agriculture (USDA) – Wildlife Services
    f. Flight safety officer

a. If operationally possible, what is the safest strategy pilots should adopt to mitigate the probability and severity of a bird strike while flying in the bird-rich zone? Why?

b. Why should you report bird strikes? How can you report a bird strike?

c. What aeronautical resources can pilots use, during their decision-making process, to gather information about the presence of birds at and around airports?

d. In which phase(s) of flight is the risk of an aircraft accident due to bird ingestion(s) the highest? Can you explain why?

e. Why should you inform air traffic control (ATC) about the presence of birds that you observe while you are flying?

f. What wildlife mitigation techniques and guidance have you been provided during your career as a pilot? By whom?

Collegiate Aviation Review International

Volume 40 | Issue 2 Peer Reviewed Article #8

11-14-2022

Creation of Tonal and Speech Alarm Efficacy Scales for Aviation

Ryan A. Lange
Embry-Riddle Aeronautical University

Stephen Rice
Embry-Riddle Aeronautical University

Cameron M. E. Severin
Embry-Riddle Aeronautical University

Jonah S. L. Chiu
Embry-Riddle Aeronautical University

Keith J. Ruskin
Embry-Riddle Aeronautical University

Connor Rice
Embry-Riddle Aeronautical University

Alarms have been in use for decades in aviation; however, it is still the case that many alarms are sub-optimally designed and do not perform well. Some alarms are so poorly designed that they increase workload, confuse the user, and/or cause a severe loss of trust. When users are asked about alarm efficacy, they often say that the alarm is either good or bad. While this provides some useful subjective information, we would argue that a quantitative scale offers more value. Using a consensus research method to ensure construct validity, we solicited 2362 participants across a four-phased, one-year study in the development of a Tonal Alarm Efficacy Scale and a Speech Alarm Efficacy Scale. A factor analysis using principal components and varimax rotation provided strong evidence of validity, while Cronbach's Alpha and Guttman's Split Half tests were used to ensure high consistency and reliability, respectively. Follow-up analyses highlight the sensitivity of the scales. These types of quantitative scales can provide a means for users, designers, engineers, and human factors experts to communicate in a common language to design more effective alarms for our society. The present study attempts to fill a gap in the current literature by providing Tonal and Speech Alarm Efficacy Scales for use applications in aviation.

Recommended Citation:
Lange, R. A., Rice, S., Severin, C. M. E., Chiu, J. S. L., Ruskin, K. J. & Rice, C. (2022). Creation of tonal and speech alarm efficacy scales. *Collegiate Aviation Review International*, 40(2), 132-145. Retrieved from http://ojs.library.okstate.edu/osu/index.php/CARI/article/view/8709/8435

Introduction

Alarms are a critical component of modern automation. They allow users of complex systems to focus on primary tasks rather than monitoring a multitude of dynamic information sources (Parasuraman et al., 2000). Many devices, including those operated by general users with no formal training, use alarms to indicate system status. For example, an automobile may sound a chime to indicate that the door is ajar or that the driver's seatbelt is not fastened, or the TCAS alarm in an aircraft may emit a vocalization indicating an imminent traffic conflict. Because of their ubiquity and the increasing complexity of modern technology, many people rely on these signals to maintain situational awareness. Poor design may impair an alarm's ability to attract the user's attention (International Organization for Standardization, 2003). Poor design may also contribute to false or nuisance alarms, which reduce the user's response to a signal (Breznitz, 1982; Dixon, Wickens & McCarley, 2007; Rice, 2009; Wickens & Dixon, 2007). Allowing users to participate in the design process may improve overall system performance (Nielsen & Levy, 1994). Therefore, the question is: Can we develop scales to evaluate the efficacy of tonal and speech-based alarms? This study aims to develop two coexistent, independent scales of efficacy that allow users to evaluate both tonal and speech alarms, creating more effective signals for aviation applications and a variety of other domains.

Literature Review

Under Parasuraman et al.'s (2000) model for levels of automation, tonal alarms can be classified as Stage 2 automation because alarm systems monitor multiple information sources within a complex system and alert the user to potential issues. Speech alarms might be classified as Stage 3 automation, as they provide verbal guidance or instructions.

85% of people in modern industrialized societies use an alarm clock to wake them (Roenneberg, 2012). The smoke detector is another commonly-used alarm that reduces the risk of fatalities by 55% (National Fire Protection Agency, 2021). These alarms are so common that we often fail to appreciate their design. Many industries—including aviation, healthcare, and nuclear power generation—employ safety-critical processes that require operators to maintain a precise mental model of system function (Carroll & Olson, 1988). These industries rely on many alarms to report system status information. The number of alarms to which users are exposed, combined with the high rate of false and nuisance alarms in some settings, suggests that a deeper understanding of their purpose and design is required (Ruskin & Hueske-Kraus, 2015).

Healthcare professionals, for example, must observe and comprehend a constant flow of data that reflects their patients' conditions. Medical equipment is designed with a comprehensive system of alarms, which notify the relevant personnel of changes in patient status hundreds of times per day (Lewandowska et al., 2020). The design of these alarms may have detrimental effects. Nurses are often subject to alarm fatigue and nuisance alarms (Ruskin & Hueske-Kraus, 2015), which can negatively impact patient safety. In a 2016 survey of clinical care providers, 30% of hospital medical staff indicated that their healthcare institution experienced adverse patient events or outcomes related to clinical alarms (Clark, 2016). In these clinical settings, an average of 150-400 alarms are generated per patient per shift, which comprises 35% of the working time of an ICU nurse (Lewandowska et al., 2020, Li et al., 2018). Recognizing an alarm, identifying its source, and interpreting its meaning require additional cognitive effort from users, especially when alarms have not been adequately designed (Ruskin & Hueske-Kraus, 2015).

In aviation, alarm-related incidents, particularly those involving the Minimum Safe Altitude Warning (MSAW), have been reported in air traffic control (Ruskin et al., 2021). These incidents are also associated with the Ground Proximity Warning System (GPWS) and the Traffic Collision Avoidance System (TCAS) (Bliss et al., 1999). Alarm fatigue may be attributable to the acoustic properties of the alarm (Edworthy, 2013). The perceived urgency of some alarms may also be difficult to judge based on their design, preventing the user from correctly responding to the alarm (Arrabito et al., 2004; Burt et al., 1995). Accurately measuring an alarm's efficacy early in the design process is a crucial factor in developing safety-critical systems.

Several measures have been used to evaluate the application of an alarm within a system. Jian et al. (2000) developed a scale that measures the level of trust a user places in an automated system. Singh et al. (1993) developed a scale that indicates the potential for complacency with an automated system by measuring attitudes toward commonly encountered automated devices. Arrabito et al. (2004) and Burt et al. (1995) have described the measurement of perceived alarm urgency using a Likert scale, which is a quick and effective way to determine perceived alarm urgency. This method also provides a way to evaluate and compare candidate alarms. Some studies have evaluated the perception of an alarm's acceptability with a single-item rating (Taylor & Wogalter, 2012). Although this may be an indicator of efficacy, an alarm that users perceive to be acceptable may not be effective for a given application. Further, many measures of existing alarm systems are focused on the metrics of alarm system performance and not the measurement of alarm system efficacy (Dorgo et al., 2021).

The Current Study

While these scales are useful for their intended purposes, they do not rate the overall efficacy of an alarm or breakdown where the problems may occur during design. A validated Likert-type scale remains to be developed that allows users to determine an alarm's perceived efficacy based on its inherent qualities. The value of single-factor rating systems is that they are quick and easy to administer. Such single-factor instruments can be delivered to operators who rely on a given alarm in the form of a survey and make it possible to easily evaluate the perceived efficacy of the alarm. The survey can be administered by various personnel, including engineers, system designers, alarm researchers, and even people without prior experience in alarm design or management. The results of these Likert-type scales are easily interpreted by the person responsible for administering and evaluating the survey, allowing the administrator to make quantitative decisions about alarms.

Prior research has shown that poor alarm design can negatively impact operator performance across many industries, including aviation. While scales have been created to measure other qualities affecting alarm performance, such as system trust and perceived urgency, none have been developed that measure perceived alarm efficacy. We have filled this research gap by developing two scales that may capture the perceived efficacy of both new and existing tonal and/or speech alarms from the users' perspective. In order to develop these scales with construct validity, we used a consensus research method (Hinkin, 1998) across a four-phased study. Participants were recruited to identify potential items for inclusion. They then narrowed those items down to a final scale. Validity, consistency, and reliability were tested using Factor Analysis, Cronbach's Alpha, and Guttman's Split Half, respectively (see Table 1 and Table 2).

Methodology – Tonal Alarm Efficacy Scale

The study was conducted with ethics approval from the university review board prior to participant recruitment.

Scale Building

The Tonal Alarm Efficacy Scale was developed in four phases: 1) item generation, 2) nominal pairing of the items, 3) Likert scale pairing, and 4) factor analysis and sensitivity test. The methods employed in this study were guided by Hinkin's (1998) framework for scale development. Similar approaches have been used to create scales in prior studies (Rice et al., 2014; Rice et al., 2015).

Phase 1: Item Generation

The overall goal of Phase 1 was to identify potential items (the written options that respondents can select from when they provide answers to questions in a survey) for integration into the final scale. This goal was accomplished by recruiting participants to complete an internet-based survey. In addition, other items were added to the list through a literature review and eight subject matter experts (SMEs) who represented human factors professionals, spaceflight engineers, airline pilots, and an anesthesiologist. Collecting inputs from SMEs has been shown to be an effective way of contributing content validity to research (Burns & Grove, 1993, p. 343).

Participants. Two hundred and two participants (94 Female, 105 Male, and 3 Other) were recruited via a convenience sample using Amazon's ® Mechanical Turk ® (MTurk) program—a crowdsourcing marketplace that allows people to participate in studies for monetary compensation. This data collection method has proven to be reliable and better represents the general population compared to laboratory data (Buhrmester et al., 2011; Thomas & Clifford, 2017). All participants were at least eighteen years of age, with a mean age of 39.92 (*SD* = 13.14) years. The participants were not screened for any specific background or technical proficiency for this phase or any other phase. Participants were paid USD $0.25 upon completion of this phase. A sample size of 200 was deemed necessary to complete the task and generate as many potential items as possible.

Procedure, Materials, and Stimuli. Participants gave their consent electronically via Google Forms ®. Following this, participants were presented with a simplified definition of what constitutes an alarm as follows:

> *An alarm is any type of auditory or visual cue that lets people know there is an ongoing danger that requires immediate action. Some alarms convey information using only tones (Tonal Alarms). Some alarms may include speech components to convey additional information (Speech Alarms).*

Participants were then presented with the following instructions:

> *In the context of the design of an alarm system, please enter 5 words or phrases that you feel are strongly relevant to the concept of a successful TONAL ALARM system design. In other words, what words or phrases would describe the qualities of a TONAL ALARM system that uses tones to convey information? For example, the alarm should be "loud," "urgent," etc. Each answer should include a word or a one-sentence phrase.*

After answering these questions, participants responded to demographic questions. Lastly, participants were compensated and thanked for their assistance.

Results. After completing the survey and consulting with the SMEs and research team, 392 unique items were generated for tonal alarms and 531 items for speech alarms. The research team reviewed the data to ensure that duplicates were either combined or removed from the list. Longer phrases were shortened to their most basic form when possible (e.g., "easy to understand" became "understandable"). In addition, researchers verified that all phrases were grammatically correct. The post-truncation item list contained 126 items for tonal alarms and 147 items for speech alarms.

Phase 2: Nominal Pairing

The goal of Phase 2 was to further refine the item list. In this phase, participants were asked to judge each term based on its relevance to the topic of "tonal alarms."

Participants. Two hundred and six participants (84 Female, 119 Male, and 3 Other) were recruited from MTurk. The average age was 41.74 (*SD* = 13.56) years. Participants were paid USD $0.50 upon completion of this phase. A sample size of 200 was deemed necessary to complete the task and maintain consensus.

Procedure, Materials, and Stimuli. The items collected in Phase 1 were presented to each participant, who rated each item on its relevance to "tonal alarms." Participants were able to rank each term as 'relevant,' 'not relevant,' or 'I don't know.'

Results. A minimum relevancy score of 70% was used to determine if an item would be included in Phase 3. Thirty-seven items met or exceeded the criteria.

Phase 3: Likert Scale Pairing

The goal of Phase 3 was to more accurately determine which items were relevant to a tonal alarm efficacy scale. In this phase, participants read through the 37 items from Phase 2 and rated them on the following scale: 0 (Not at all related to tonal alarm qualities), +1 (Slightly related to tonal alarm qualities), +2(Somewhat related to tonal alarm qualities), +3(Quite related to tonal alarm qualities), +4 (Extremely related to tonal alarm qualities).

Participants. Two hundred and thirty-nine participants (79 Female, 159 Male, and 1 Other) were recruited through a convenience sample using MTurk. The average age was 35.59 (*SD* = 10.00) years. Participants were paid USD $0.30 upon completion of this phase. A sample size of 200 was deemed necessary to complete the task and maintain consensus.

Results. Six terms met the inclusion criteria and were included in the final Tonal Alarm Efficacy scale: *effective*, *attention-grabbing*, *audible*, *loud*, *useful*, and *identifiable*.

Phase 4: Factor Analysis and Sensitivity Test

In this phase, the final scale of six items was assessed for validity, reliability, and sensitivity. Participants were given a hypothetical scenario in which an alarm would be necessary and were instructed to listen to one of three alarm stimuli. The stimuli were created with Audacity® audio editing software version 3.1.3 and uploaded onto YouTube®. Following this, participants responded to the scenario using the Tonal Alarm Efficacy scale (see Appendix A).

Participants. Six hundred and nine participants (347 Males and 262 Females) were recruited using a convenience sample from MTurk. The average age was 38.90 (*SD* = 11.29) years. Participants were paid USD $0.20 upon completion of this phase. A sample size of at least 600 was deemed necessary for conducting the factor analysis. As a general rule, the sample size required for factor analysis is roughly ten times the number of variables you are testing (Comrey & Lee, 1992).

Procedure, Materials, and Stimuli. Participants were randomly divided into three groups based on alarm quality and presented with the following scenario:

> *Imagine a scenario in which you are in a building. A serious event has occurred and the building's evacuation alarm sounds. The following alarm is sounded and you must respond accordingly. Play the video to hear the alarm, and then answer the following question.*

Each group listened to one of three alarm stimuli that were designed to be either “low-quality” (n = 243), “mid-quality” (n = 229), or “high-quality” (n = 157). If needed, participants could listen to the alarm as many times as they wished. Participants then responded to the Tonal Alarm Efficacy scale (see Appendix A).

Results. A factor analysis using the principal components and varimax rotation showed that all items strongly loaded on a single factor for each group, with 59-62% of the variance explained for each model (see Table 1). The scale also had a very high level of internal consistency, as determined by Cronbach's alpha values of 86-88%. Guttman's Split Half tests indicated very high reliability with results between 88-90%.

Table 1
Statistical Analysis Results

Analyses performed	Low-Quality Alarm	Mid-Quality Alarm	High-Quality Alarm
Variance explained	0.62	0.59	0.62
Cronbach's alpha	0.88	0.86	0.88
Guttman Split Half Test	0.90	0.88	0.89

Before analysis, scores of the scale were averaged to create a single "alarm efficacy" score for each participant. The scores for the three groups were then compared using a one-way ANOVA with an LSD post hoc test. Alarm efficacy was statistically significantly different between the alarm quality groups $F(2, 606) = 30.63$, $p < .001$, $\eta^2 p = 0.092$. Alarm efficacy increased from the Low-Quality alarm group ($M = 0.49$, $SD = 0.89$) to the Mid-Quality alarm group ($M = 1.03$, $SD = 0.71$) and to the High-Quality alarm group ($M = 0.98$, $SD = 0.74$), indicating that the scale is sensitive to different levels of efficacy.

Methodology – Speech Alarm Efficacy Scale

Participants. One thousand one hundred and six participants (501 Female, 596 Male, and 9 Other) were recruited from MTurk. The average age was 38.32 ($SD = 11.43$) years. Participants were compensated in the same manner as Phases 1, 2, 3, and 4 of the Tonal Alarm Efficacy Scale development process (USD $0.25, $0.50, $0.30, and $0.20, respectively). Sample size determinations were also identical to the development of the Tonal Alarm Efficacy Scale.

Procedure, Materials, and Stimuli. The Speech Alarm Scale was developed exactly like the previous Tonal Alarm Scale. In Phase 1, 531 unique items were generated. These were pared down to 32 items in Phase 2 and seven items in Phase 3: *effective, attention-grabbing, simple, audible, clear, reliable, and understandable* (see Appendix B).

Results. A factor analysis using the principal components and varimax rotation revealed that all items strongly loaded on a single factor for each group, with 59-62% of the variance explained for each model (see Table 2). The scale also had a very high level of internal consistency as determined by Cronbach's alpha values of 86-88%. Guttman's Split Half tests indicated very high reliability, with results between 90-91%.

Table 2
Statistical Analysis Results

Analyses performed	Low-Quality Alarm	Mid-Quality Alarm	High-Quality Alarm
Variance explained	0.62	0.59	0.58
Cronbach's alpha	0.88	0.86	0.86
Guttman Split Half Test	0.91	0.90	0.90

Before analysis, scores of the scale were averaged to create a single "alarm efficacy" score for each participant. The scores for the three groups were then compared using a one-way ANOVA with an LSD post hoc test. Alarm Efficacy was statistically significantly different between the alarm quality groups $F(2, 665) = 14.70$, $p < .001$, $\eta^2 p = 0.042$. Alarm efficacy increased from the Low-Quality alarm group ($M = 0.36$, $SD = 0.94$), to the Mid-Quality alarm group ($M = 0.47$, $SD = 0.85$), and to the to the High-Quality alarm group ($M = 0.79$, $SD = 0.78$), indicating that the scale is sensitive to different levels of efficacy.

Discussion and Conclusions

The purpose of the current study was to create and validate two scales that can be used to study users' perceived efficacy of alarms. To achieve this, we conducted a four-phased study to ensure construct validity. Participants, including eight SMEs, were recruited to identify potential items for inclusion and then to narrow those items down into a final scale. Validity, consistency, and reliability were tested using Factor Analysis, Cronbach's Alpha, and Guttman's Split Half, respectively.

The use of a consensus methodology for scale development contributed to the construct validity of the two scales (Hinkin, 1998). Factor analysis further supported the validity of the scales, showing that all the items for both the Tonal Alarm Efficacy Scale and the Speech Alarm Efficacy Scale contribute to a single factor: "tonal alarm efficacy" and "speech alarm efficacy", respectively. The results of the Cronbach's Alpha calculation indicated a high level of internal consistency among the items. The reliability of the scales was tested using Guttman's Split Half tests. Unlike the test-retest method, this method only requires one administration of the scales, reducing the need for participants and allowing for easier administration over the internet.

Other scales developed prior to this study focused primarily on evaluating one or more factors that may contribute to an alarm's efficacy, such as user trust or complacency (Jian et al., 2000; Singh et al., 1993). Many studies have focused on alarm system performance metrics (Dorgo et al., 2021), but none have sought to fully capture a user's perceived efficacy of an alarm. The alarm efficacy scales will facilitate the integration of user input in the development of alarm systems. Ultimately, this will allow designs to be driven by human factors principles (Nielsen & Levy, 1994) while providing quantitative data to support design decisions.

The alarm efficacy scales are highly suitable for use in the aviation industry. They are easy to administer and implement into the design process and capable of producing actionable results. Due to their generalizable nature, the scales are also suitable for use in a variety of other industries. For example, an administrator working on the design of the flight deck for a new aircraft might begin by collecting or creating the alarms to be evaluated. Next, the administrator recruits a sample of participants who represent the alarm system's intended user: pilots. The pilots are invited to participate in the evaluation individually. Each pilot is first given information and context about the system that contains the candidate alarms. The pilot must have as much context as possible about the situation in which the alarm is utilized so that they may form an adequate opinion about the alarm's properties. The pilot is also briefed about the purpose of the scale and how to fill it out.

Next, the pilot listens to one of the candidate alarms and evaluates it by filling out the Tonal or Speech scale. The pilot is then presented with the next candidate alarm, and the process is repeated until the pilot has listened to and rated each of the candidate alarms. The administrator can then compare the scores and choose the best alarm.

Additionally, a designer could compare a current alarm with an improved version. For example, the designer might use the Speech Alarm Efficacy scale to have users rate a newly-designed speech alarm being considered for use in an updated version of the TCAS alarm in an aircraft. Based on the ratings of this alarm, the designer can implement specific changes to its design, creating an iterative process where user feedback is solicited whenever new changes are made to the candidate alarm. Ultimately, a highly refined version of the candidate alarm would be selected for use in the finished product.

This four-phased study was conducted to develop and test a scale for measuring the perceived efficacy of tonal and speech alarms. These scales are intended for use by equipment designers, manufacturers, and users. Both scales are short and easy to administer, making them ideal for use in the iterative design processes used to create and evaluate alarms over the phases of a design project. The Tonal and Speech Alarm Efficacy Scales will give users a voice in the design of alarms, ultimately improving the safety of those who rely on them.

Limitations

The current study has several limitations. First, participants were recruited using convenience sampling techniques. Since participants reported that they live in the United States, these results may be limited in perspective to western ideologies, leaving room for future studies to enhance the generalizability of the scales to international audiences. Additionally, control of the environment in which participants responded to auditory stimuli was not regulated. The volume of the stimuli, repetitions, the presence or absence of background noise and headphones, and the possibility that participants were distracted could not be controlled using the current study design. Finally, participants were not screened for hearing deficiencies prior to listening to the alarm samples, which may have impaired their ability to judge the efficacy of the alarms.

References

Arrabito, G. R., Mondor, T. A., & Kent, K. J. (2004). Judging the urgency of non-verbal auditory alarms: a case study. *Journal of Ergonomics, 47(*8*)*, 821-840. https://doi-org.ezproxy.libproxy.db.erau.edu/10.1080/00140130420001932828

Bliss, J. P., Freeland, M. J., & Millard, J. C. (1999). Alarm related incidents in aviation: A survey of the aviation safety reporting system database. *Proceedings of the Human Factors and Ergonomics Society. Annual Meeting, 1*, 6.

Breznitz, S. (2013). *Cry wolf: The psychology of false alarms*. Psychology Press.

Buhrmester, Kwang, T., & Gosling, S. D. (2011). Amazon's Mechanical Turk: A New Source of Inexpensive, Yet High-Quality, Data? *Perspectives on Psychological Science, 6*(1), 3–5. https://doi.org/10.1177/1745691610393980

Burns, N., & Grove, S. K. (1993). The practice of nursing research: Conduct, critique & utilization (2nd ed). Sanders.

Burt, J. L., Bartolome, D. S., Burdette, D. W., & Comstock, J. R. (1995). A psychophysiological evaluation of the perceived urgency of auditory warning signals. *Ergonomics*, *38*(11), 2327-2340. doi:10.1080/00140139508925271

Carroll, J. M., & Olson, J. R. (1988). Mental models in human-computer interaction. *Handbook of human-computer interaction*, 45-65.

Clark, T. (2016, June 5). *HTF Update: 2016 National Clinical Alarm Survey Results*. AAMI Conference and Expo, Tampa, FL. http://thehtf.org/documents/2016%20National%20Clinical%20Alarms%20Survey%20Results.pdf

Comrey, A. L., & Lee, H. B. (1992). *A first course in factor analysis* (2nd ed). L. Erlbaum Associates.

Dixon, S. R., Wickens, C. D., & McCarley, J. S. (2007). On the independence of compliance and reliance: Are automation false alarms worse than misses? *Human Factors, 49*(4), 564-572.

Dorgo, G., Tandari, F., Szabó, T., Palazoglu, A., & Abonyi, J. (2021). Quality vs. quantity of alarm messages - How to measure the performance of an alarm system. *Chemical Engineering Research and Design*, *173*, 63-80.

Edworthy, J. (2013). Medical audible alarms: A review. *Journal of the American Medical Informatics Association: JAMIA, 20*(3), 584-589. doi:10.1136/amiajnl-2012-001061

Hinkin, T. R. (1998). A brief tutorial on the development of measures for use in survey questionnaires. *Organizational Research Methods, 1*(1), 104–121. http://dx.doi.org/10.1177/109442819800100106

International Organization for Standardization (2003). Ergonomics — Danger signals for public and work areas — Auditory danger signals (ISO Standard No. 7731:2003). https://www.iso.org/standard/33590.html

Jian, J. Y., Bisantz, A., & Drury, C. (2000). Foundations for an Empirically Determined Scale of Trust in Automated Systems. *International Journal of Cognitive Ergonomics*, *4*, 53–71. https://doi.org/10.1207/S15327566IJCE0401_04

Lewandowska, K., Weisbrot, M., Cieloszyk, A., Mędrzycka-Dąbrowska, W., Krupa, S., & Ozga, D. (2020). Impact of Alarm Fatigue on the Work of Nurses in an Intensive Care Environment-A Systematic Review. *International journal of environmental research and public health*, *17*(22), 8409. https://doi.org/10.3390/ijerph17228409

Li, T., Matsushima, M., Timpson, W., Young, S., Miedema, D., Gupta, M., & Heldt, T. (2018). Epidemiology of patient monitoring alarms in the neonatal intensive care unit. Journal of Perinatology, 38(8), 1030-1038.

National Fire Protection Agency (2021). Smoke Alarms in US Home Fires. Retrieved from https://www.nfpa.org/News-and-Research/Data-research-and-tools/Detection-and-Signaling/Smoke-Alarms-in-US-Home-Fires

Nielsen, J., and Levy, J. (1994). Measuring usability — preference vs. performance. Communications of the ACM 37, 4 (April), 66–75.

Parasuraman, R., Sheridan, T. B., & Wickens, C. D. (2000). A model for types and levels of human interaction with automation. *IEEE Transactions on systems, man, and cybernetics-Part A: Systems and Humans*, *30*(3), 286-297. https://www.ida.liu.se/~769A09/Literature/Automation/Parasuraman,%20Sheridan,%20Wickens_2000.pdf

Rice, S. (2009). Examining single and multiple-process theories of trust in automation. *Journal of General Psychology, 136*(3), 303-319.

Rice, S. C., Mehta, R., Winter, S., & Oyman, K. (2015). A trustworthiness of commercial airline pilots (T-CAP) scale for American consumers. *Journal of Aviation Technology and Engineering*, *4*(2), 55.

Rice, S., Mehta, R., Steelman, L. A., & Winter, S. R. (2014). A trustworthiness of commercial airline pilots (T-CAP) scale for Indian consumers. *International Journal of Aviation, Aeronautics, and Aerospace*, *1*(3), 3.

Roenneberg, Till. *Internal Time : Chronotypes, Social Jet Lag, and Why You're So Tired*, Harvard University Press, 2012. *ProQuest Ebook Central*, http://ebookcentral.proquest.com/lib/erau/detail.action?docID=3301120.

Ruskin, K. J., & Hueske-Kraus, D. (2015). Alarm fatigue: impacts on patient safety. *Current Opinion in Anesthesiology*, *28*(6), 685-690.

Singh, I. L. Molloy, R., & Parasuraman, R. (1993). Automation-induced "complacency": Development of the Complacency-Potential Rating Scale. *The International Journal of Aviation Psychology*, *3*(2), 111-122

Taylor, J. R. I., & Wogalter, M. S. (2012). Acceptability of Evacuation Instruction Fire Warnings. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *56*(1), 1753–1757. https://doi.org/10.1177/1071181312561352

Thomas, K. A., & Clifford, S. (2017). Validity and mechanical turk: An assessment of exclusion methods and interactive experiments. *Computers in Human Behavior*, *77*, 184-197. https://doi.org/10.1016/j.chb.2017.08.038

Wickens, C. D., & Dixon, S. R. (2007). The benefits of imperfect diagnostic automation: A synthesis of the literature. *Theoretical Issues in Ergonomics Science*, *8*(3), 201-212.

Appendix A –Tonal Alarm Efficacy Scale

Please respond how strongly you agree or disagree with the following statements.

	Strongly Disagree	Disagree	Neither Agree nor Disagree	Agree	Strongly Agree
1. The alarm is Effective.	-2	-1	0	1	2
2. The alarm is Attention-grabbing.	-2	-1	0	1	2
3. The alarm is Audible.	-2	-1	0	1	2
4. The alarm is Loud.	-2	-1	0	1	2
5. The alarm is Useful.	-2	-1	0	1	2
6. The alarm is Identifiable.	-2	-1	0	1	2

The final alarm efficacy score will be the average of the six responses.

Appendix B - Speech Alarm Efficacy Scale

Please respond how strongly you agree or disagree with the following statements.

	Strongly Disagree	**Disagree**	**Neither Agree nor Disagree**	**Agree**	**Strongly Agree**
1. The alarm is Effective.	-2	-1	0	1	2
2. The alarm is Attention-grabbing.	-2	-1	0	1	2
3. The alarm is Simple.	-2	-1	0	1	2
4. The alarm is Audible.	-2	-1	0	1	2
5. The alarm is Clear.	-2	-1	0	1	2
6. The alarm is Reliable.	-2	-1	0	1	2
7. The alarm is Understandable	-2	-1	0	1	2

The final alarm efficacy score will be the average of the seven responses.

Collegiate Aviation Review International

Volume 40 | Issue 2 Peer Reviewed Article #9

12-18-2022

Predicting Pilot-in-training Success and Persistence in a United States University

Maria Chaparro Osman
Florida Institute of Technology

Vivek Sharma
Florida Institute of Technology

Cherrise Ficke
Florida Institute of Technology

Rian Mehta
Florida Institute of Technology

Brooke Wheeler
Florida Institute of Technology

Meredith Carroll
Florida Institute of Technology

Commercial pilot numbers have been on a decline since 2009, while in parallel, pilot demands continue to rise. In order to close the gap, airlines and companies need a steady stream of pilots-in-training who have successfully graduated. However, collegiate flight training programs have had issues with low retention and low success rates of pilots-in-training. The need to understand how to retain students within collegiate flight programs and increase success rates is vital to fill the gaps present within the aviation industry. Past studies have investigated predictive factors for pilots-in-training graduation persistence and the time it takes to graduate yielding similar findings. Many factors have been identified as related to pilot-in-training success and persistence, the most common being high school GPA and cost. However, suggestions have been presented as variables of interest to examine for future studies, such as the effect of different types of flight postponements. The research study conducted expanded the knowledge regarding variables that predict or contribute to pilot-in-training success in a collegiate aviation flight degree program, examining graduation persistence and time to graduate. An archival data of 262 pilot-in-training students were used to explore the relationship between the 19 predictors, graduation persistence, and time to graduate. Several variables had significant relationships with both Time to Graduate and Persistence Before Dropout, including Age, Number of Transfer Credits, Class Load, and Pass Rates in Aeronautics Classes. Additionally, Time to Graduate had a significant relationship with Academic Success in Aeronautics Classes and Maintenance Postponements per Semester. Persistence Before Dropout had a significant relationship with Instructor Changes, Instructor Postponements, and Weather Postponements.

Recommended Citation:
Osman, M.C., Sharma, V., Ficke, C., Mehta, R., Wheeler, B., & Carroll, M. (2022). Predicting pilot-in-training success and persistence in a united states university. *Collegiate Aviation Review International*, 40(2), 146-164. Retrieved from http://ojs.library.okstate.edu/osu/index.php/CARI/article/view/8809/8436

Introduction

Background

The aviation industry continues to grow as air travel becomes safer and more affordable (Airlines for America, 2019). Airline passenger traffic and revenue passenger miles (RPM) have increased significantly, with 2018 seeing an increase of over 109 million RPM in the United States between 2015 and 2018 (Airlines for America, 2018; Lutte, 2019). Unfortunately, the increases seen for passengers has not mirrored in commercial pilots, as the pilot numbers that would be needed to fulfill passengers' needs are not there (Caraway, 2020; Federal Aviation Administration, 2019; Klapper & Ruff-Stahl, 2019) in parallel to the rising increase in pilot demand (CAE, 2017; Lutte, 2018). Higgins et al. (2013) projected that U.S. airlines could face a pilot shortage between 2013 and 2031 as this could go as high as 35,000 pilots. Although the COVID-19 pandemic decreased pilot-in-training numbers for a period of time, a survey administered to aviation flight schools around the world found that pilot-in-training numbers are back to 94 percent of their pre-COVID-19 activity levels and still rising (Flight Logger, 2021; Redbird Flight Simulations, 2021). The aviation industry is expected to rebound from setbacks caused by COVID-19, and the substantial need for pilots is expected to continue (Redbird Flight Simulations, 2021). In order to close the gap between passenger demand and the shortage of commercial pilots, airlines and companies need a steady stream of successfully graduating pilots-in-training. However, collegiate flight training programs have had issues with low retention, and low success rates of pilots-in-training (De Montalk, 2000; Leonard, 2018).

There is a need to understand how to retain students within collegiate flight programs and increase success rates to fill the gaps present within the aviation industry. A limited number of studies in the past have investigated predictive variables related to pilots-in-training success such as flight costs, class load, and academic success in flight. Patterns have emerged; however, more research is needed examining these variables together to further understand their relationship to persistence before dropout and the time it takes to graduate. The studies associated with predictor variables of pilot-in-training graduation success are presented in the following paragraphs. Leonard (2018) examined student success using Time to Graduate, measured as whether they graduated in 48 months or longer, and cumulative GPA, measured as the overall GPA at the end of the four years in a collegiate aviation program. Leonard framed the research question based on Astin's Input, Environment, and Output model (1970, 1984). A multiple regression was used to analyze 3,100 students registered in different aviation majors at the University of North Dakota during the 2016-17 academic year. The results of the study indicated that there was a significant relationship between credit loads, which were defined as academic intensity and degree attainment. Students who registered for higher credit loads were more likely to graduate in 48 months compared to students who took fewer credits each year on average. Additionally, working as a flight instructor increased the likelihood of graduating in 48 months. Leonard proposed that students who take more credit hours and work as flight instructors are spending more time surrounded by students who are going through similar experiences and can benefit from constant exposure.

McFarland (2017) conducted a correlation analysis to investigate the relationship between pre-entry and flight training attributes of students as predictors of success in a Flight Training program. Three construct groups were used: (a) an academic construct measured using test scores; (b) a cognitive construct using cognitive tests; and (c) a performance construct consisting of flight course grades and days to course completion. A sample of 242 students who completed a private pilot flight course was utilized. Results were consistent with collegiate math and/or physics scores, high school GPA, and ACT scores positively correlating with the successful completion of the private pilot course. Performance in flight courses was also found to be significantly correlated with successful completion of the private pilot course. When examining the relationship between the cognitive constructs and performance in flight courses, no significant relationship was found. McFarland suggested that reviewing performance in flight courses during the beginning of the program would help identify at-risk students early enough to intervene. The findings of the study also suggested that a decrease in the student-to-instructor ratio facilitated a better student retention rate in the flight program.

Bjerke and Healy (2010) examined pre-entry attributes as predictors of students' persistence and academic success in a Flight Training program. Persistence was measured as students who continued from the spring into the subsequent fall semester, and academic success was measured using GPA from the term. The study used archival data from 390 full-time students enrolled in a public, four-year, research-intensive university with commercial aviation as their declared major. The sample consisted of two cohorts, one beginning in the fall of 2006, and the other beginning in the fall of 2007. The following predictor variables were used in the study: Age, sex, ethnicity, high school GPA, ACT score, math ACT score, verbal ACT score, family's gross income, father's education level, mother's education level, and admitted credit hours. A multiple regression analysis was used to investigate the relationship between pre-entry attributes and the two criterion variables. Pre-entry attributes accounted for 9.6% of the variance in the students' persistence, with high school GPA as the most significant predictor. Variables such as age ($r = .04$), admitted credit hours ($r = .10$), and family's gross income ($r = .42$) were positively related to persistence. The results also indicated that the pre-entry attributes accounted for 32.3% of the variance in academic success. High school GPA ($r = .47$), admitted credit hours ($r = .05$), and ACT math scores ($r = .16$) had a significant and positive relationship with academic success. However, sex had no significant relationship with students' persistence and academic success. The findings suggest that older students, those who come into the program with more credit hours, and students whose families have higher gross incomes would have higher persistence in a collegiate aviation program. Additionally, those with higher high school GPAs, higher credit hours upon being admitted to the program, and higher ACT math scores were likely to have higher academic success.

The Government Accountability Office (GAO, 2018) investigated the variables contributing to the ability of 147 flight schools to produce professional pilots in the United States. GAO reviewed the following data from flight schools offering professional flight degree programs: degree completion for the 2015-16 school year, the FAA's data on flight schools' enrollment, the certified flight instructor data, the number of pilots in different categories (e.g., flight instructor and recreational), and new pilot certificates issued. Thirty-five representatives from flight schools, airports, airline personnel, FAA officials, and flight instructors were interviewed. The results of the study indicated that the high cost of training and low flight-instructor retention were the key challenges affecting flight schools' ability to produce professional pilots. Additionally, it was determined that both scheduled and unscheduled aircraft maintenance could also reduce the ability of flight schools to train flight students.

The Airline Owners' Pilot Association (AOPA, 2010) conducted a survey to investigate the variables that could improve student success rates in flight programs. The study used a mixed-methods approach. First, a qualitative survey asking participants to rate various attributes, such as instructor effectiveness, scheduling, and costs, was disseminated to create a list of variables that would lead to optimal Flight Training performance and increase the student success rate. Second, quantitative research was conducted to identify the perceptions of student pilots and flight instructors regarding the list of variables collected by the qualitative survey. The study used a sample of 1,000 students and flight instructors (750 students, 250 flight instructors). A list of 67 variables (attributes) identified through the qualitative analysis was divided into four groups based on the quantitative analysis: education quality, customer focus, community, and information sharing. Education quality consisted of instructor support, instructor effectiveness, and organized lessons. Customer focus consisted of variables such as flight costs, scheduling, and aircraft quality. The community consisted of certifications for student achievement and aviation environment. Information sharing consisted of variables such as success rates, estimated time, costs, and instructor experience level. The resulting participant responses indicated that improving flight instructor and student relationships would increase flight program experience and student success rates. Additionally, the results indicated that scheduling lessons in advance and being flexible would help in the Flight Training experience.

This extant research, though limited, suggests that there is a range of different predictor variables that likely impact Persistence Before Dropout and graduation success. It is also apparent from these findings that additional research is needed to confirm which variables are the most influential. The previous studies guided the research effort in many ways. The results of Leonard's (2018) study indicated that there was a significant relationship between higher credit loads and degree attainment as well as working as a flight instructor and graduating in 48 months. Therefore, the average class load per semester was examined in the current study to further explore this relationship. McFarland (2017) found a positive correlation between standardized test scores and completion of the private pilot course. Performance in flight courses was also positively correlated with the completion of the private pilot course. Therefore, the current study included a standardized test score variable and flight training performance variable. Bjerke & Healey (2010) found that age, admitted credit hours, and family's gross income were positively related to persistence. Further, high school GPA, admitted credit hours, and ACT math scores had a significant and positive relationship with academic success. Therefore, age, socioeconomic status, transfer credits, and academic variables such as GPA in flight and aeronautics courses were included. The results of the research conducted by GAO suggested that training costs, low flight-instructor retention, and aircraft maintenance were key in affecting pilot-in-training numbers. Therefore, the current study included variables for flight cost, instructor changes, and maintenance postponements. Finally, research by the AOPA suggested that improved flight instructor and student relationships could increase student success rates. The results also indicated that scheduling lessons in advance and flexibility would help in the flight training experience. Therefore, the current study examined instructor changes, student postponements, and flight instructor postponements.

Methodology

The purpose of the current research effort was to expand the knowledge regarding variables that predict or contribute to pilot-in-training success in a collegiate aviation flight-degree program by examining pilots-in-training persistence before dropping out and the time it takes to graduate. An ex post facto design was conducted using 19 predictor variables in five different functional sets. Set A = individual difference variables, Set B = involvement variables, Set C = achievement variables, Set D = instructor interaction variables, and Set E = flight postponement variables (see Table 1 for all variables within each set). The variables in each set were found to have a relationship with the dependent variables of Time to Graduate or Persistence Before Dropout for pilots-in-training.

Research Questions

There were two primary research questions (RQ):

> RQ 1: What is the relationship between variables related to student individual differences, involvement, achievement, instructor interactions, flight postponements, and time to graduate from a collegiate aviation training program?
>
> RQ 2: What is the relationship between variables related to student individual differences, student involvement, student achievement, instructor interactions, flight postponements, and persistence before dropout from a collegiate aviation training program?

Participants

Archival data were obtained from the Office of Institutional Research (OIR) at a Part 141 Collegiate program in the Southeastern U.S. for all flight students majoring in Aeronautical Science and Aviation Management with flights from 2010 until 2019. For the purpose of the current study, only student data between 2010 and 2016 were included in the dataset. This range ensured that all the students who were included in the analysis had been enrolled in the degree program for at least four years. The first regression was conducted for Time to Graduate and included a sample of n = 141 flight students. The second regression was conducted with Persistence Before Dropout as the dependent variable and included a sample of n = 121 flight students.

Power Analysis

An a priori power analysis was run using G* Power (Faul et al., 2007) for an F-test, using a multiple regression based on the five sets of variables: Power analysis parameters included an $\alpha = 0.05$, with a medium effect size of .15 (Cohen et al., 2003) and 80% power. Based on the

G*Power results, it was determined that a minimum sample size of 163 participants was necessary to detect a medium effect size in the population. Unfortunately, our sample did not meet the minimum sample size requirements indicated by G*Power. However, similar correlational studies have reported a coefficient of determination (R^2) of more than .30 (Leonard et al., 2018), corresponding to a large effect, given that we had an R^2 of .52 for Persistence Before Dropout and an R^2 of .60 for Time to Graduate, we were comfortable moving forward with the analysis.

Experimental Design and Variables

The current study used a predictive correlational design and two hierarchical regressions to investigate the relationship between the five sets of predictor variables and their relationship with two different criterion variables: Persistence Before Dropout and Time to Graduate. The Persistence Before Dropout variable was calculated by taking the number of semesters a student completed out of the total expected academic semesters in the major (i.e., eight semesters). The Time to Graduate variable was measured as the number of semesters it took a student to graduate. The independent variables were selected and categorized into five sets based on the literature and anecdotal reports from flight instructors. Categorical variables were coded using a dummy coding strategy. Variables are presented in Table 1 and described in the following section.

Set A included eight individual difference variables. There were five continuous and three categorical variables. Age was a continuous variable, and socioeconomic status was a continuous variable measured using the financial aid received by the student in total dollars. Cognitive ability was measured as a continuous variable using a composite of two standardized test scores (SAT and ACT). Major was a categorical variable, students whose major was Aeronautical Science with Flight were represented as 0, and Aviation Management with Flight was represented as 1. The flight costs variable was measured using the total flight fees paid by the student during their time enrolled in the university and was presented in two ways as the variability in the number of semesters for the Persistence Before Dropout variable had a high variability with respect to the number of semesters students persisted. Specifically, we chose to categorize the flight costs based on the flight line's minimum threshold costs for each academic year. For the first regression in which Time to Graduate was the dependent variable, Flight Cost was a continuous variable. For the second regression in which Persistence Before Dropout was the dependent variable, Flight Cost was a dummy coded categorical variable with three categories (high, medium, and low) to allow high and medium Flight Costs to be compared to Low Flight Costs. High, Medium, and Low Flight Costs were calculated using the university flight line fees and policies posted on their website. Once calculated, two standard deviations below this value and two standard deviations above this value were used to create a minimum and maximum cost a student could have during their first, second, third, and fourth year in flight school. For example, the estimated flight cost for the first year was $23,759. Therefore, any student who paid between $18,000 to $21,000 fell into the Low Flight Cost category. Students who paid between $21,000 and $24,000 fell under the Medium Flight Cost category, and any price above $24,000 was considered to be High Flight Cost. Certification was measured by whether the student came in with their private pilot license or not. Finally, the Transfer Credits variable was measured using the total number of credits transferred into the program by the student.

The variable in Set B was measured in the following manner: Class Load is a continuous variable that was measured by calculating the average class load per semester. The variables in Set C were represented in the following manner: Pass Rate was a continuous variable represented by the percentage of aeronautics classes passed based on the total number of classes taken. Academic Success in Aero Classes was a continuous variable represented by the GPA of all aeronautics classes combined. Academic Success in Flight Classes was a continuous variable represented by the GPA of all flight classes combined.

The variables in Set D were represented in the following manner: Instructor changes were a continuous variable, representing the total number of different flight instructors the students trained with while obtaining their certificates. Flight Training performance was a continuous variable measured using the percentage of passed check rides based on total check rides. Set E consisted of five flight postponement variables, including Instructor Postponements, Student Postponements, Weather Postponements, Maintenance Postponements, and Other Postponements. Each postponement variable was a continuous variable in which the number of relevant postponements was averaged per semester.

Analysis

Preliminary Analysis

Outliers were removed based on studying jackknife distances and visual inspections of scatterplots for each variable. One outlier from each regression was removed based on these analyses. The final sample size for the Time to Graduate model was 140 participants, and the final sample size for the Persistence Before Dropout model was 120 participants.

Table 1
Independent Variable Sets and Criterion Variables in the Current Study

Sets/Independent Variables	Measure Description
Set A = Individual Difference Variables	
X_1 = Sex	X_1 is a categorical variable, self-reported sex
X_2 = Age	X_2 is a continuous variable, student age upon entry
X_3 = Socioeconomic Status	X_3 is a continuous variable, the dollars of financial aid
X_4 = Cognitive Ability	X_4 is a continuous variable, SAT/ACT composite of scores
X_5 = Major	X_5 is a categorical variable, one of two Aeronautics majors
X_6 = Flight Costs	X_6 is a continuous variable, total flight fees paid for Time to Graduate
	X_6 is a categorical variable with three levels (high, medium, and low) for Persistence Before Dropout
X_7 = Certification	X_7 is a categorical variable, flight certifications upon entering program
X_8 = Transfer Credits	X_8 is a continuous variable, number of transfer credits
Set B = Involvement Variables	
X_9= Class Load	X_9 is a continuous variable, average class load per semester
Set C = Achievement Variables	
X_{10} = Pass Rates	X_{10} is a continuous variable, percentage of Aero classes passed
X_{11} = Academic Success in Aero Classes	X_{11} is a continuous variable, Aeronautics class GPA
X_{12} = Academic Success in Flight Classes	X_{12} is a continuous variable, Flight class GPA
Set D = Instructor Interaction Variables	
X_{13} = Instructor Changes	X_{13} is a continuous variable, average number of instructor changes per semester
X_{14} = Flight Training Performance	X_{14} is a continuous variable, percentage of passed stage checks
Set E = Flight Postponement Variables	
X_{15} = Instructor Postponements	X_{15} is a continuous variable, average number of instructor postponements per semester
X_{16} =Student Postponements	X_{16} is a continuous variable, average number of student postponements per semester
X_{17} = Weather Postponements	X_{17} is a continuous variable, average number of weather postponements per semester
X_{18} = Maintenance Postponements	X_{18} is a continuous variable, average number of maintenance postponements per semester
X_{19} = Other Postponements	X_{19} is a continuous variable, average number of other postponements per semester
Dependent Variables	
Y_1 = Time to Graduate	Y_1 is a continuous variable, the number of semesters it took for a student to graduate
Y_2 = Persistence Before Dropout	Y_2 is a continuous variable, the percentage calculated based on the number of semesters in a flight major out of the total academic semesters (i.e., eight)

Regression Assumptions

Next, both models were tested for the six regression assumptions. Assumption 1, the assumption of linearity, was not met by the Time to Graduate model. To resolve this issue, we built the bi-variate scatter plots for each independent and dependent variable. We found that Class Load was not linearly related to the Time to Graduate model. Based on Cohen et al. (2003), squared and cubic models were created for the Class Load variable to facilitate the interpretation of a non-linear slope. The persistence model met the linearity assumption. For assumption 2, the correct specification of the IVs, we chose a *p-value* threshold of 0.2, which means all the variables that had a *p*-value less than 0.2 were included in the final model. For the Time to Graduate regression, only seven out of the 17 variables met these criteria and were utilized in the primary analysis, including Age, Major, Transfer Credits, Class Load, Pass Rates, Academic Success in Aero, and Average Number of Maintenance Related Postponements per Semester. For the Persistence Before Dropout regression, nine out of the 17 variables met these criteria and were utilized in the primary analysis, including Age, High Flight Costs, Low Flight Costs, Class Load, Pass Rates, Instructor Changes, Instructor Postponements, Student Postponements, and Weather Postponements. Both the models met assumption 3, perfect reliability; assumption 4, homoscedasticity of residuals; assumption 5, independence of residuals; and assumption 6, normality of residuals.

Primary Analysis

Two hierarchical regressions were conducted with Persistence Before Dropout and Time to Graduate as the criterion variables and 19 predictor variables divided into five sets as described in Table 1. Based on the literature, a prioritized set entry order of A-B-C-D-E was utilized. In Table 2, we present the final model for Time to Graduate, which had an overall R^2 = .60, meaning 60% of the variance in Time to Graduate is collectively explained by individual difference variables, involvement variables, achievement variables, and flight postponement variables. The overall model was significant, $F\ (7, 132) = 27.71$, $p < .001$, and included significant independent variables of Age, Transfer Credits, Academic Success in Aero Classes, and Maintenance Postponements. When individual difference variables were added into the model, the overall $R^2 = 0.19$, with Age and Transfer Credits contributing a significant amount of variance, indicating individual difference variables uniquely contributed 19% of the variance in Time to Graduate. When involvement variables were added into the model, the overall $R^2 = 0.55$, indicating involvement variables uniquely contribute 36% of the variance in Time to Graduate with one significant variable, Class Load. When achievement variables were added into the model, the overall R^2= 0.59, indicating that achievement variables uniquely contributed approximately 4% of the variance in Time to Graduate, including one significant variable, Academic Success in Aero Classes. When flight postponement variables were added to the model, the overall $R^2 = 0.60$. Average Flight Postponement uniquely contributed about 1% variance in the Time to Graduate, including one significant variable, Maintenance Postponements per Semester.

Table 2
Final Hierarchical Model of Time to Graduate

Predictor	Model 1B	Model 2B	Model 3B	Model 4B	
				B	**95% CI**
Constant	4.7***	15.9***	26.9***	26.91**	[12.41,41.42]
X2 = Age	0.14**	0.08*	0.09*	0.08*	[0.001,0.153]
X5 = Major Code	0.15	0.52*	0.39	0.40	[-0.04, 0.83]
X8 = Number of Transfer Credits	0.02**	0.01*	0.01*	0.01*	[0.001, 0.02]
X9 = Class Loads		-0.63***	-0.65**	-0.66*	[-0.77, -0.54]
X10 = Pass Rates in Aero Classes			-13.52	-13.17	[-28.05,1.70]
X11 = Academic Success in Aero Classes			0.73**	0.74**	[0.23, 1.24]
X18 = Maintenance Postponements				0.74*	[0.04, 1.44]
Statistical Results					
R^2	0.19***	0.55***	0.59***	0.60***	
F	10.52	42.17	30.88	27.71	
Delta R^2		0.36***	0.04***	0.01	
Delta F	0.19***	0.55***		0.60	

Note. *Indicates *p*-value is significant at 0.05. ** Indicates *p*-value is significant at 0.01. *** Indicates *p*-value is significant at 0.001.

In Table 3, we present the final model for Persistence Before Dropout, which had an overall R^2 = 0.52, meaning 52% of the variance in Persistence Before Dropout is collectively explained by individual difference variables, involvement variables, achievement variables, instructor interaction variables, and flight postponement variables. The overall model was significant, F (9, 111) = 16.63, p < .001, and included significant independent variables of Age, Transfer Credits, Academic Success in Aero, and Maintenance Postponements per Semester. When individual difference variables were added into the model, the overall R^2 = 0.08., with Age being a significant variable.

Table 3
Final Hierarchical Model for Persistence Before Dropout

Predictor	Model 1B	Model 2B	Model 3B	Model 4B	Model 5B	
					B	95% CI
Constant	8.55***	17.8***	14.99***	16.79***	18.01***	[11.22, 24.80]
X_2 = Age	-0.16**	-0.23	-0.22***	-0.17**	-0.205***	[-0.31, -0.10]
X_6^a = High Flight Costs	0.74	0.90	0.316	2.1**	3.81***	[2.27, 5.35]
X_6^b = Number of Transfer Credits	1.1	1.12	0.646	1.55*	2.28*	[1.08, 3.5]
X_9 = Class Load		-0.49*	-0.49*	-0.53**	-0.53*	[-0.85, -0.27]
X_{10} = Pass Rates in Aero Classes			3.29*	3.03**	3.15**	[1.14, 5.16]
X_{11} = Instructor Changes				-1.72***	-1.31***	[-1.18, -0.80]
X_{15} = Instructor Postponements					-0.26*	[-0.49, -0.03]
X_{16} = Student Postponements					0.15	[-0.06, 0.38]
X_{17} = Weather Postponements					-0.32	[-0.52, -0.12]
Statistical Results						
R^2	0.08*	0.12**	0.17***	0.45***	0.52***	
F	3.38	3.92	4.65	15.06	16.63	
Delta R^2		0.04***	0.5***	0.28***	0.07***	
Delta F						

Note. * Indicates p-value is significant at 0.05. ** Indicates p-value is significant at 0.01. *** Indicates p-value is significant at 0.001.

When involvement variables were added into the model, the overall $R^2 = 0.12$, indicating involvement variables uniquely contributed 4% of the variance in Persistence Before Dropout with one significant variable, Class Load. When achievement variables were added to the model, the overall $R^2 = 0.17$, indicating that achievement variables uniquely contributed about 5% of the variance in Persistence Before Dropout with one significant variable, Academic Success in Aero Classes. When instructor interaction variables were added into the model, the overall $R^2 = 0.45$, indicating that instructor interaction variables uniquely contributed about 28% of the variance in Persistence Before Dropout with one significant variable, Instructor Changes. When flight postponement variables were added to the model, the overall $R^2 = 0.52$. Flight postponement variables uniquely contributed about 7% variance in Persistence Before Dropout, with two significant variables, Instructor Postponements and Student Postponements per Semester.

Discussion

The results of the current study support previous findings and provide support for variables related to individual differences, involvement, achievement, instructor interaction, and flight postponements being predictive of pilot-in-training success. Several variables had significant relationships with Time to Graduate and Persistence Before Dropout, including Age, Number of Transfer Credits, Class Load, and Pass Rates in Aeronautics Classes. Additionally, Time to Graduate had a significant relationship with Academic Success in Aeronautics Classes and Maintenance Postponements per Semester. Persistence Before Dropout had a significant relationship with Instructor Changes, Instructor Postponements, and Weather Postponements. These relationships will be examined in detail below.

Individual Difference

A significant and positive relationship was found between Age and Time to Graduate. Specifically, for every 10-year increase in age, on average, Time to Graduate increased by one semester. Similarly, for every 10-year increase in age, on average, Persistence Before Dropout decreased by half a semester. This finding suggests that as age increases, pilots-in-training are less likely to graduate within eight semesters and are less likely to persist. This finding may be due to changes in responsibilities that accompany increases in age, such as family and work responsibilities which limit the ability to dedicate time to studies and can result in part-time enrollment (Bjerke & Healey, 2010; Shapiro et al., 2016). Alternatively, this may be influenced by the smaller number of students in the sample who were older than the traditional collegiate age range (i.e., 18-22 years old) or differences in the types of students who are non-traditional (i.e., delayed college, changed careers, returned to college after military service). With this in mind, it may be beneficial to make older students more aware of the program requirements and time expenditure necessary to complete the program within eight years. This finding is not consistent with some of the literature (e.g., Leonard, 2018; Waldman & Avolio, 1986). Leonard (2018) found no relationship between age and graduating within 48 months. The difference may be due to a lack of variability in the age of their sample size as their mean was 18, and their standard deviation was 0.5, whereas the current sample had a mean of 25 and a standard deviation of 3.2 for the Time to Graduate regression and 5.2 for the Persistence Before Dropout regression. Bjerke and Healy (2010) had similar findings, in which age was positively correlated to persistence from year one to year two. However, just as in Leonard (2018), the variables were examined in a different light that is they did not look at it across semesters.

A significant relationship was also found between High Flight Costs and Persistence Before Dropout. Students with High Flight Costs persisted four semesters longer than students with Low Flight Costs. High flight costs were, costs two standard deviations higher than the average cost that a student could pay based on flight line fees and policies. Low flight costs were two standard deviations lower than the average flight cost that a student could pay based on flight line fees and policies. Students with Medium Flight Costs (i.e., the average flight cost based on flight line fees and policies) persisted for two semesters longer than students with Low Flight Costs. This suggests that the students who paid high flight costs persisted longer than the students who paid medium and low flight costs. This finding may point to the sunken cost fallacy (Friedman et al., 2007), as students are focusing on the time and money already invested and do not want it to go to waste, so they persist. Another reason may have to do with the population from which the sample was derived. The study university has a large international student population, and anecdotal reports from advisors and certified flight instructors point to the fact that pressure from families back home may encourage students to persist with the flight major that they initially started with, and this explanation has also been supported by previous scientific literature (Andrade, 2008; Kwai, 2010).

A positive and significant relationship was also found between the number of transfer credits a student transferred in when they started the program and the Time to Graduate. For every additional 100 credits transferred into the university, Time to Graduate increased by one semester. This finding is not in line with extant research. Bjerk and Healy (2010) found that the higher the transfer credits, the higher the academic success, which does not align with the current findings. A potential explanation for this is that many individuals transfer in credits from a different major, but not all credits can be utilized towards the flight degree. Therefore, additional credits need to be taken into account based on the number of credits counting towards the flight degree program.

There were several individual difference variables that were not found to have significant relationships. No relationship was found between Sex and either Time to Graduate or Persistence Before Dropout. Studies examining sex and pilots-in-training have found similar results (Bjerke & Healy, 2010; Leonard, 2018); however, these results have not been found in non-aviation contexts (Peltier et al., 2000; Trippi & Baker, 1989;). This may be due to the similarities in personality between males and females in the aviation industry, a consistent finding that is not replicated in the general population and is an influencing factor in aviation training outcomes (Chaparro et al., 2020). No significant relationship was found between Major and the two dependent variables. This may be due to the inclusion of students from the two flight majors, aeronautical science and aviation management, being too similar. Interestingly, no relationship was found between Flight Costs and Time to Graduate, which again may point to the university at which the study was conducted, as instructors and advisors stated that many of the international students were not concerned with the money aspect as long as they graduated from the program. This is an interesting finding from the perspective of reverse causation, where the longer it takes one to graduate because of the flight, the more they pay toward tuition. This finding points to what is commonly referred to as the "sunk cost fallacy," as students have invested so much time and money that they are unwilling to stop in the hopes that they can turn things around (Siegel, 2011).

No relationship was found between Transfer Credits a student held and Persistence Before Dropout: therefore, this may not be a factor that causes students to persist in a program. A possible explanation for this finding could be that when a student transfers in with more credits, they have less time to finish the degree and, therefore, may not encounter problems of dropout before finishing the degree. For example, if a student is transferring with 60 credits, the student only needs, on average, two years to finish the remaining 60 credits to graduate; therefore, both the short duration and less number of credits will not significantly impact students' Persistence Before Dropout. Based on the finding, we recommend that it might be beneficial if flight programs orient the students transferring in with a previous degree with oversight to help them graduate on time. Those students should be made aware that their Time to Graduate may take around a semester more, depending on their situation.

Involvement

When examining involvement, specifically the Class Load variable, a polynomial relationship resulted between Time to Graduate and Class Load. Time to Graduate increased until approximately 12 credit hours per semester, where it then decreased until 15 credit hours per semester; after 15 credit hours, the Time to Graduate increased again. The findings of the current study were consistent with extant research (Attewell & Monaghan, 2016; Attewell et al., 2012; Huntington-Klein & Gill, 2020) that suggests there is an optimal class load, under which students may not be taking enough classes to stay on track, and over which student load may be overtaxing and extend their time due potentially to failing classes. When examining Class Load and Persistence Before Dropout, a negative relationship was found. For every six additional credits per semester, Persistence Before Dropout decreased, on average, by one semester. This may point to the threshold for too many classes before an individual's persistence is affected. Based on the findings, we recommend that universities encourage students to take between 12 and 15 credits during the spring and fall semesters.

Achievement variables

A positive relationship between Pass Rates and Persistence Before Dropout was found. For every 30% increase in Pass Rates, on average, Persistence Before Dropout increased by one semester. The findings were consistent with the study conducted by Clery and Topper (2008) that illuminated how students who do well are likely to persist in the program longer than those with low pass rates. A positive relationship also existed between student Academic Success in Aeronautics Classes and both Time to Graduate and Persistence Before Dropout, indicating that for every 1-point increase in GPA, Time to Graduate increases by approximately one semester. For every 30% increase in aeronautics class Pass Rates, on average, Persistence Before Dropout increased by one semester. These findings may seem unintuitive; however, many high-performing students pursue one or more minors. Pursuing a minor requires students to take additional courses, such as graduate courses, which could impact Time to Graduate (Clery & Topper, 2008).

No relationship was found between Pass Rates and Time to Graduate, potentially due to the fact that high performers often take on a minor or enroll in additional classes, which may add on time, whereas those who are doing well and not taking extra classes may actually graduate early, making that relationship difficult to discern.

Instructor Interactions

A negative relationship was found between Instructor Changes and Persistence Before Dropout. For every two instructor changes per semester, Persistence Before Dropout decreased by one semester. A plausible explanation for this finding could be the CFIs being international students. International students, after obtaining a flight instructor license, work towards their commercial license, which requires 1,500 hours of flying time. Being a CFI, once they finish 1,500 hours of flying, they may leave the CFI job to start working as an airline pilot. In the context of this study, students who work with a particular CFI for a few semesters may lose interest if their CFI leaves the job. This may be due to the rate at which instructor changes tend to occur by either student's progress through check rides versus an issue with an instructor. Unlike other university programs, the classroom for a pilot is more high-risk and requires a feeling of comfort with their instructors, a constant change not related to check rides may not allow for that trust to build, leading to decreased persistence.

No relationship was found between Flight Training performance and the two dependent variables. This may be due to many factors, such as the fact that different instructors grade students in a different manner, and each student is graded by multiple instructors. Additionally, students can receive different flight certifications, leading to differing levels and opportunities for performance results. No relationship was found between Instructor Changes per semester and Time to Graduate, potentially due to the fact that instructor changes happen throughout the program in part because instructors have different certification levels, which means that they can only teach certain flight courses. Thus, it would not likely affect the time it takes an individual to graduate. However, this is not the case with persistence.

Flight Postponements

Weather and Instructor postponements had a negative relationship with Persistence Before Dropout. That is, for every three additional Weather and Instructor Postponements per Semester, Persistence Before Dropout decreases by one semester. A plausible reason for this finding could be the students' locus of control and motivation level. This may point to a difference in those who persist at a motivational or locus of control level. For example, students with an external locus of control may get discouraged, as they cannot control an instructor or weather postponing their flight, in turn leading to a lack of motivation to continue (Dille & Mezack, 1991). Based on the findings, it may be beneficial for flight programs to attempt to limit instructor changes when possible or put programs in place designed to create effective matches between instructors and students.

A positive relationship was found between Maintenance Postponements and Time to Graduate. For every ten additional maintenance-related postponements per semester, on average, Time to Graduate increases by one semester. A possible explanation for this finding could be that, depending on the type of maintenance, a certain plane may be out of commission for a long period of time, leading to fewer options in the fleet for all students to choose from. This could lead to students' slow progress in flight classes and, therefore, delay the time to graduation. Moreover, Maintenance postponements are completely out of the students' control and may not only put them behind but may also discourage them in their progress.

No significant relationship was found between Student Postponements and Persistence Before Dropout; however, there may be practically significant findings. No relationship was found between Instructor Postponements and Time to Graduate. This may be due to the fact that those who graduate are very intrinsically motivated and are not affected by instructors postponing. No relationship was found between Student Postponements per semester and Time to Graduate; Weather Postponements and Time to Graduate; Maintenance postponements per semester and persistence; or Other Postponements and the two dependent variables. The lack of relationship between other postponements and the DVs may be due to the small number of "other" postponements (i.e., a small sample size). Additionally, the fact that the rest of the postponement variables may have covered the aspects that more commonly affect students. Furthermore, anecdotal reports by instructors support that students tend to. Based on the above findings, it might be beneficial for flight schools to put programs in place to cope with postponements. Students who begin acquiring large numbers of postponements should be made aware of potential impacts and provided assistance for making up lost ground and avoiding future delays.

Limitations, Delimitations, and Future Research

These findings should be interpreted with caution given several limitations of the study. First, we were not able to control how any of the data was collected by the university. This led to missing data. For example, both socioeconomic status and cognitive ability were removed from the analysis as there was too much missing data (i.e., 30% of the data were missing). This data may have added more insight. However, excluding variables in which 30% of the data is missing aides in unbiased results (Cohen et al., 2003). Future archival research should try to include an alternative, reliable and valid measure to capture any missing data of this kind. Furthermore, due to the recency of the creation of the Aviation Human Factors major, and lower enrollment in Aviation Meteorology, we were not able to include all flight majors. Therefore, future research should include students from other flight majors such as aviation human factors and meteorology. Future studies should attempt to replicate the study in different universities while collecting data related to student ethnicity to obtain a better understanding of the potential influences of international student pressures. Additionally, since we were using archival data, we were not able to collect all of the data we would have ideally used, such as information regarding student scholarships and grants, which could help augment the flight cost information. Future studies should also attempt to collect additional data, such as personality, motivation, locus of control, ethnicity, and scholarships. Finally, the sample data collected for the current study was a non-probability-based convenience sampling strategy, as all the archival data which was available from 2010 - 2016 was obtained from a single private university limiting the generalizability of the results. Future research should include data from more flight schools.

Conclusion

The current study established a prediction model with six sets of predictor variables and two dependent variables which were Time to Graduation and Persistence Before Dropout among pilot-in-training students in a Southeastern college in the U.S. The results indicate that Age, Number of Transfer Credit, Class Loads, and Pass Rates in Aero Classes are significant predictors of Time to Graduation and Persistence Before Dropout. In addition to that, Major Code, Academic Success in Aero Classes, and Maintenance Postponements are significant predictors of Time to Graduation. High Flight Costs, Instructor Changes, Student Postponements, and Weather Postponements are significant predictors of Persistence Before Dropout. We recommend older students – make aware of the program requirements and time expenditure. Class load – Students should be encouraged to take between 12 and 15 credits. Transfer students - may require more oversight or help to graduate on time. Postponements – Minimize postponements; make students aware of potential impacts and provide assistance in making up lost ground. Instructor Changes – Programs should attempt to limit instructors' changes.

References

Airlines for America. (2018, October 24). Industry Review. Retrieved from Allocating Capital to Benefit Customers, Employees, and Investors: https://airlines.org/wpuploads/2017/03/A4A-Industry-Review-5.pdf

Airlines for America. (2019). U.S. Airline Traffic and Capacity. Retrieved from U.S. Airlines (Passenger and Cargo): https://www.airlines.org/dataset/annual-results-u-s-airlines-2/

AOPA. (2010, October). The Flight Training Experience. Retrieved from A Survey of Students,

Andrade, M. S. (2008). International student persistence at a faith-based institution.

Astin, A. W. (1970). The methodology of research on college impact, part one. *Sociology of education*, 223-254.

Astin, A. (1984). Student Involvement: *A developmental Theory for higher Education. Journal of College Student Personnel, 25*(4), 297–308.

Attewell, P., & Monaghan, D. (2016). How many credits should an undergraduate take?. *Research in Higher Education, 57*(6), 682-713.

Attewell, P., Heil, S., & Reisel, L. (2012). What is academic momentum? And does it matter?. *Educational Evaluation and Policy Analysis, 3*4(1), 27-44.

Bjerke, E., & Healy, M. (2010). Predicting student persistence: Pre-entry attributes that lead to success in a collegiate flight program. *The Collegiate Aviation Review International, 28*(1).

Caraway, C. L. (2020). A looming pilot shortage: It is time to revisit regulations. International Journal of Aviation, Aeronautics, and Aerospace, 7(2), 3. Chicago

Chaparro, M. E., Carroll, M., & Malmquist, S. (2020). Personality Trends in the Pilot Population. *The Collegiate Aviation Review International, 38*(2).

Clery, S., & Topper, A. (2008). Students Earning Zero Credits. Data Notes. Volume 3, Number 5, September/October 2008. *Achieving the Dream*.

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). Applied multiple regression/correlation analysis for the behavioral sciences (3rd ed.). Lawrence Erlbaum Associates Publishers.

De Montalk, R. J. (2000). Anxiety as a factor in student pilot performance in a university aviation degree programme. Massey University. Massey University: Doctoral Dissertation.

Dille, B., and M. Mezack. 1991. Identifying predictors of high risk among community college telecourse students. *American Journal of Distance Education 5*(1): 24–35.

FAA. (2019). Civil airmen statistics. Retrieved from Not APA https://www.faa.gov/data_research/aviation_data_statistics/civil_airmen_statistics/

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.

Flight Logger. (2021, May 11). Flight School Management Software. Retrieved from https://flightlogger.net/

Friedman, D., Pommerenke, K., Lukose, R., Milam, G., & Huberman, B. A. (2007). Searching for the sunk cost fallacy. *Experimental Economics*, *10*(1), 79–104.

GAO. (2018). Collegiate Aviation Schools. Stakeholders Views on Challenges for Initial Pilot Training Programs. GAO –18-403. https://www.gao.gov/assets/gao-18-403.pdf

Higgins, J., Lovelace, K., Bjerke, E., Lounsberry, N., Lutte, R., Friedenzohn, D., . . . Craig, P. (2013). *An investigation of the United States airline pilot labor supply*. Retrieved from https://www.researchgate.net/publication/249315130_An_Investigation_of_the_United_States_Airline_Pilot_Labor_Supply

Huntington-Klein, N., & Gill, A. (2021). Semester course load and student performance. *Research in higher education, 62*(5), 623-650.

Klapper, E. S., & Ruff-Stahl, H. J. K. (2019). Effects of the pilot shortage on the regional airline industry: A 2023 Forecast. International Journal of Aviation, Aeronautics, and Aerospace, 6(3), 2.

Kwai, C. K. C. (2010). Model of international student persistence: Factors influencing retention of international undergraduate students at two public statewide four-year university systems. (Doctoral dissertation, University of Minnesota).

Leonard, A. P. (2018). *The Impact of Pre-entry Attributes and College Experiences on Degree Attainment for Students in a Collegiate Flight Program*. The University of North Dakota

Lutte, B. (2018). Pilot supply at the regional airlines: Airline response to the changing environment and the impact on pilot hiring. Journal of Aviation/Aerospace Education & Research, 27(1), 1-22.

Lutte, R. K. (2019). Women in aviation: A workforce report.

McFarland, M. R. (2017). *Student Pilot Aptitude as an Indicator of Success in a Part 141 Collegiate Flight Training Program* (Doctoral dissertation, Kent State University).

Redbird Flight Simulations. (2021, March 17). *State of Flight Training: 2021 Survey and Report*

[PowerPoint Slides]. Download. https://simulators.redbirdflight.com/state-of-flight-training

Peltier, G. L., Laden, R., & Matranga, M. (2000). Student persistence in college: A review of research. Journal of College Student Retention: Research, Theory & Practice, 1(4), 357-375.

Shapiro, D., Dundar, A., Wakhungu, P.K., Yuan, X., Nathan, A, & Hwang, Y. (2016, September). Time to Degree: A National View of the Time Enrolled and Elapsed for Associate and Bachelor's Degree Earners (Signature Report No. 11). Herndon, VA: National Student Clearinghouse Research Center.

Siegel, M. J. (2011). Reimagining the retention problem: Moving our thinking from end-product to by-product. *About Campus*, *15*(6), 8-18.

Trippi, J. F., & Baker, S. B. (1989). Student and residential correlates of black student grade performance and persistence at a predominantly white university campus. *Journal of College Student Development*, *30*, 135-143.

Waldman, D. A., & Avolio, B. J. (1986). A meta-analysis of age differences in job performance. *Journal of applied psychology, 71*(1), 33.

Wilson, N., Bjerke, E., & Martin, L. (2015). Aviation living learning community: Impacts on student success. *Collegiate Aviation Review*, *33*(1), 29–43.

Collegiate Aviation Review International

Volume 40 | Issue 2 Peer Reviewed Article #10

12-19-2022

Development of Critical Thinking Skills in Collegiate Aviation Programs

Irene Miller
Southern Illinois University

Thomas Long
Central Washington University

Critical thinking requires an individual to gather and interpret data, develop conclusions based upon relevant findings, and implement the best solution. The dynamic aviation industry requires these skills of its pilots, maintainers, and managers for companies to remain successful. Collegiate aviation programs need to teach critical thinking and cognitive skills to allow students entering the workforce to become these successful aviation managers, maintainers, and pilots. The School of Aviation at Southern Illinois University conducted a research study to inform the development of pedagogical techniques for promoting critical thinking skills in the classroom and determine their efficacy. The research study used critical thinking skills assessments, in a pre and posttest format, to determine if students experienced an increase in their ability to think critically. The findings indicate there was no statistically significant increase in students' total scores on the critical thinking assessments. However, these findings support the development of best practices which can be adopted and refined by collegiate aviation programs.

Recommended Citation:
Miller, I. & Long, T. (2022). Development of critical thinking skills in collegiate aviation programs. *Collegiate Aviation Review International*, 40(2), 165-183. Retrieved from http://ojs.library.okstate.edu/osu/index.php/CARI/article/view/9380/8437

Introduction

Statement of the Problem

Aviation-related companies find critical thinking to be a valuable skill for employees (Bhatti, 2020). The aviation industry is dynamic and requires aviation managers to adapt quickly to remain successful. Aviation managers need to, based upon trends, determine the direction the industry is heading to allow their organizations to effectively evolve and prosper. Previous research indicates the need for aviation managers to possess strong critical thinking skills but does not provide direction on how to teach critical thinking skills to students in collegiate aviation programs (Davitch & Folker, 2017). Critical thinking skills increase an individual's ability to solve problems, assess risk and make informed decisions (Bhatti, 2020).

Purpose of the Study

The purpose of the research study is to develop pedagogical techniques to promote critical thinking skills in the classroom and determine efficacy. To achieve this objective, students must be able to define critical thinking, identify specific critical thinking skills, and practice the application of these skills in the classroom. To this end, pedagogical techniques need to be developed to teach critical thinking skills to students enrolled in collegiate aviation programs.

Research Questions

1. Does the comparison of the scores on the pre-course and post-course critical thinking skills assessments indicate whether the students increased their critical thinking skills through pedagogical techniques?

2. What specific portions of the pre-course and post-course critical thinking skills assessments showed variations based upon a comparison of the scores?

Significance of the Study

Critical thinking is the ability to objectively evaluate information and develop a logical conclusion (Faruki, 2019). Aviation-related organizations find critical thinking to be a valuable skill for employees. Employees that possess critical thinking skills can solve problems, identify logical connections between ideas, evaluate arguments, and use logical reasoning to reach conclusions (Bhatti, 2020). The development of critical thinking skills amongst collegiate aviation students will likely assist these students as they pursue their careers in the aviation industry.

Assumptions & Limitations

For this study, the following limitations and assumptions are proposed:

1. This study is limited based upon the voluntary participation of two groups of collegiate aviation students asked to participate in this research study. One group consisted of collegiate aviation management students enrolled in AVM 374 General Aviation Operations course at Southern Illinois University during the fall 2021 semester. The next group consisted of collegiate aviation management students enrolled in the same course at Southern Illinois University during the spring 2022 semester. Both groups of students are pursuing a Bachelor of Science in Aviation Management at SIU. There was a low participation rate during the spring 2022 semester, so the data for both groups of students were combined for analysis.

2. This study is limited by certain uncontrollable variables that cannot be accounted for, such as student motivation, commitment, and academic aptitude.

3. It is assumed the students answered the pre-course survey and critical thinking assessment questions honestly to the best of their knowledge.

4. The same critical thinking assessment, hereafter referred to as the pretest and posttest, respectively, was used at the beginning and end of the course. The students were not able to review the pretest until they completed the posttest at the end of the semester. There was a 15-week span between the pretest and posttest given to the students. Therefore, it is assumed that students would not remember the questions when completing the posttest.

5. The critical thinking assessment (pretest and posttest) used for this research study was gathered from the Watson-Glaser critical thinking appraisal practice test, which provides validity to the critical thinking assessment used for this research study.

Literature Review

In the classroom, students absorb information in order to pass an examination, but once the examination is completed, they forget what they learned. Terenzini et al. (1993) found, “A relatively large body of evidence suggests that much of the factual material students learn during college may have a relatively short shelf-life after they leave (McLeish, 1968; Gustav, 1969; Blunt & Blizard, 1975; Brethower, 1977). It is critical to teach students how to retain that information. Several authors have defined critical thinking in various ways. According to Richard Paul, critical thinking is not a new concept. The terms “critical” and “thinking” have Greek and Latin roots. Critical thinking is described as “thinking that strives to know by judging with discernment and using standards and tests as a means” (CriticalThinkingOrg, 2015a, 7:12). Others define critical thinking differently; Richard Ennis (1993) states, "Critical thinking is reasonable reflective thinking focused on deciding what to think or do" (p. 180). Critical thinking is the ability to use intelligence, knowledge, and skills to question and carefully explore situations and arrive at thoughtful conclusions based on evidence and reason (Neck, 2016). Critical thinking is also defined as a mode of reasoning in which one improves the quality of thought by skillfully analyzing, assessing, and reconstructing their thought processes (Davitch & Folker, 2017). Although there are numerous interpretations of critical thinking, a worldwide multidisciplinary panel defined it as "the process of conscious, self-regulating judgment" (Bycio & Allen, 2009, p. 2).

Organizations require managers who can think independently, without bias or judgment, and forecast patterns of behavior and procedures. In order to make successful and deliberate decisions, they ask the appropriate questions: how and why, rather than just what (Neck, 2016). Plummer (2019) noted most companies lack an efficient method for objectively assessing critical thinking skills and are unsure how to deliver coaching to team members who need to improve their critical thinking abilities. The ability to think critically and problem-solving requires the ability to think fundamentally and systemically, particularly in the context of problem-solving (Cahyo et al., 2020). The business world requires individuals with excellent communication skills, the ability to work in teams, and problem-solving abilities. The ability to solve problems is a high-level cognitive ability (Cahyo et al., 2020). Belkin (2017) noted test results from exams administered between 2013 and 2016 indicated that the typical graduate showed little or no progress in critical thinking across four years at universities.

Companies have been using pre-employment testing to determine which applicant is a good fit for the organization. One of the skills that businesses look for in employees is the ability to identify solutions to challenges. Employers place a high value on the ability to provide a well-thought-out solution in a fair amount of time. Other organizations seek people who can perform tasks beyond reading, writing, and math, such as critical thinking and problem-solving. Companies/organizations might expect a distinct level of corporate culture if they have the ability to use critically improved thinking. In the long run, improved culture may convert into money or more income, but in the near run, it may translate into enhanced personal communications, cooperation, and collaboration (Murawski, 2014).

Management needs employees who can make decisions with minimal supervision, allowing them to make decisions on their own (“What is the RED Model,” 2019). Kuncel and Hezlett (2010) made the following observation: Numerous research studies were undertaken in both educational and employment settings to examine if cognitive assessments predict performance. The findings were inconsistent, with some research claiming that cognitive test scores are virtually perfect indicators of performance and others claiming that cognitive test scores are unrelated to academic and professional achievement. According to Belkin (2017), a PayScale Inc. poll indicated that 50% of employers say college graduates they recruit are not ready for the profession, with weak critical-reasoning abilities being the most common complaint. In fact, Belkin reports 60% of managers saying that recent graduates lack critical thinking skills (Belkin, 2017).

Ennis (1993) stated there are many critical thinking examinations; however, he found no subject-specific categorization in the collection of critical thinking tests. Existing multiple-choice tests do not appropriately examine several crucial aspects of critical thinking, such as being open-minded and cautiously reaching justified conclusions. The critical thinking essay test is an anomaly; it covers more ground than the others, but it is less secure and requires more time and money than multiple-choice examinations. Using the Watson-Glaser test, Daryl Smith (1977) discovered that increased praise, peer-to-peer connection, and student participation lead to more high-critical-thinking behaviors. Furthermore, as these processes become more prevalent, memorization habits become less prevalent. Watson and Glaser identified five critical-thinking skills: inference, assumption recognition, deduction, interpretation, and argument evaluation. The Watson-Glaser Critical Thinking Appraisal (WGCTA) proved to be the most effective tool for evaluating hypotheses and research questions (Stone, 2016). The WGCTA is designed to measure those important abilities and skills involved in critical thinking (WGCTA, 2012). Robert Ennis's view on most critical thinking tests is that they are not comprehensive, especially those that are easiest to use, such as the multiple-choice tests (Ennis, 1993). Critical thinking skills should be measured over time, just as instructor pilots conduct periodic check rides for their students (Davitch et al., 2017, p. 63). The WGCTA has been used to predict performance in a variety of educational settings (Ejiogu et al., 2006). Recent research "finds that cognitive aptitude tests, particularly critical thinking exams, ... are among the strongest and most consistent predictors of success across academic and work settings" (Neck, 2016, p. 1). Critical Thinking Competency Standards provide a framework for assessing students' critical thinking abilities. It enables administrators, teachers, and faculty at all levels to determine the extent to which students are reasoning critically within any subject or discipline (Paul & Elder, 2005).

Currently, there is a scarcity of research on methods for building important abilities in potential pilots. The business world requires individuals with excellent communication skills, the ability to work in teams, and problem-solving abilities. The ability to solve problems is a high-level cognitive ability (Cahyo et al., 2020). The link between critical thinking and problem-solving abilities is also vital for aviation mechanics in the field. Critical thinking is a talent that is frequently required in aircraft repair. This expertise is vital since aircraft technicians are often confronted with troubleshooting issues. Mechanics must draft and retain repair records, as well as document all preventative and corrective maintenance on the aircraft (Cahyo et al., 2020).

A study done by Herasymenko et al. (2019) on aviation studies found that the "education system in flying institutions did not" appear to "focus on Aviation English as a tool for developing critical thinking skills (Herasymenko et al., 2019, p. 100). These researchers emphasized the link between critical thinking abilities and linguistic ability. They acknowledge that programs focused on strengthening critical thinking can improve language abilities. Incorporating critical thinking exercises into the core of the Aviation English course can result in improved aviation safety (Herasymenko et al., 2019).

In the introduction of Richard Paul's lecture on "How to teach students to listen and read properly," listening and reading are only effective learning tools when done correctly, yet most students are deficient in elementary listening and reading skills (CriticalThinkingOrg, 2015b). The essential goal in education should be to advance students' ability to think critically (Trapasso, 2021). The ability to think critically and to solve problems leads to the ability to think fundamentally and systemically, particularly in the context of problem-solving (Cahyo et al., 2020). Adam Stone (2017) discovered the need to find instructors with good critical thinking scores to teach critical thinking, as students mirror their instructors' critical thinking scores. Stone noted research done by Lois Magnussen on a nursing program suggests those students with low scores improved to approximate the teachers' critical thinking scores, while students with previously similar scores with the instructor have remained nearly the same. Nursing students with high critical thinking scores plummeted to average scores over the course of the multiyear program (Stone, 2017). An analysis from Huber and Kuncel (2016) found nursing students simply did not improve more than their non-nursing counterparts. For other specialties, it is unknown if critical thinking grows linearly or more rapidly throughout the college years (Huber & Kuncel, 2016).

To be successful in the aviation sector, pilots, mechanics, and management must be able to think critically. Research has found students study for the exam and forget what they learned shortly after graduation. Management in the aviation industry feels new graduates lack critical thinking and problem-solving skills. Acquiring those skills is vital in higher education, yet there is evidence that some students have shown little progress in critical thinking after four years of study. To improve students' ability to think critically, it was found through research that hiring teachers with great critical thinking abilities would increase a student's critical thinking skills to the same level as the instructor's scores. A range of assessments, practice, hands-on repetition and feedback have been found to increase students' critical thinking skills. In exchange, students need to become lifelong learners and commit to the continual improvement of their critical thinking skills.

Methodology

Research Design

This study is based upon applied research and focuses on the development of critical thinking skills within the classroom and the performance of students completing critical thinking assessments used to validate an improvement in critical thinking skills.

Once data collection and analysis were complete, the researchers were able to determine if the critical thinking materials and exercises provided in class were effective at increasing students' critical thinking skills. It is assumed that higher scores on the posttest indicate an increase in participants' critical thinking skills. Only then can pedagogical techniques be developed and evaluated to teach critical thinking skills to students enrolled in collegiate aviation programs.

A mixed methodology approach was used for data collection, as the researchers applied quantitative and qualitative inquiry during the research study. There is a need for the hard data that quantitative research provides; however, the soft data that qualitative analysis provides fills the gaps left by quantitative data or provides needed context (Patton, 2015). Combining the two research methods allowed the researchers to develop accurate and comprehensive conclusions. The purpose of the mixed method approach was to determine the efficacy of pedagogical techniques to increase critical thinking skills.

Target Population and Participant Selection

This research used purposeful sampling. Purposeful sampling focuses on a smaller sample to allow for a comprehensive analysis rather than a larger sample which provides a cursory investigation (Patton, 2015). Both qualitative and quantitative inquiry was used to collect data. This study focused on depth by using a smaller sample size. The sample size was influenced by several factors, such as lack of compensation for participants and voluntary participation. There was no compensation provided for students to participate in the study, which may have discouraged participation. Also, participation was voluntary as the pretest and posttest scores were not part of students' course grades. This impacted students' commitment to complete both assessments as some students completed the pretest at the beginning of the semester but chose not to complete the posttest at the end of the semester. Despite the small sample size, sufficient data was collected, and meaningful conclusions were reached. Data were collected from two groups of students. Two groups of students were enrolled in the AVM 374: General Aviation Operations course, using a traditional 16-week format, during the Fall 2021 and Spring 2022 semesters.

Participants

The critical thinking skills of 27 students in the SIU School of Aviation were assessed. Seven students were eliminated from the study because of incomplete data (either did not take the pretest or posttest), leaving 20 participants with sufficient data for analysis. The participants were 70% male and 30% female, with a mean age of 22.6 years. 65% of participants identified as White, 15% as Hispanic, 10% as Black, and 10% as Other. The mean GPA for the participants was 3.42, with 3.2 being the mode GPA of the participants. Ten percent of the students were sophomores, and 45% were juniors and seniors.

Procedures

Eighteen (n=18) students during the fall 2021 semester and nine (n=9) students during the spring 2022 semester participated in the research study. However, only two (n=2) of the spring 2022 students completed the research study by completing the posttest leaving a total of 20 participants with sufficient data for analysis. The students completed the pretest during the first week of class and completed the posttest on the last day of class. All the students who participated during the fall 2021 semester completed the posttest, but seven of the students during the spring 2022 semester did not complete the posttest.

The critical thinking measure was administered at the beginning of the semester and at the end of the semester to explore possible changes to students' critical thinking skills as a result of participation in the course. The Watson-Glaser Critical Thinking Appraisal is the most commonly used psychometric test by organizations for the pre-selection of managers (Pearson, n.d.). Participants' critical thinking skills were assessed using the Watson-Glaser Critical Thinking Appraisal in a pretest, posttest research design. The Watson-Glaser Critical Thinking Appraisal uses the RED model, which organizes critical thinking based on the ability to recognize assumptions, evaluate arguments, and draw conclusions (Pearson, n.d.). Studies have indicated a positive correlation between those individuals that score well on the Watson-Glaser Critical Thinking Appraisal and job and course success (Pearson, n.d.).

Data Collection

The qualitative data from the research study was collected through the student pre-course survey, which provided a great deal of qualitative data about the students that could not be realized through the pretest and posttest. Specifically, the student pre-course survey collected information such as demographics, educational history, and other pertinent information. The qualitative data was analyzed to illuminate the quantitative data. Without this valuable qualitative data, the quantitative data gathered from the student pretests and posttests stands as it is yet to provide little understanding. Analyzing the quantitative data through the lens of qualitative data allowed the researchers to provide credible answers to the research questions.

Research Instruments

The student pre-course survey consisted of 10 closed and seven open-ended questions for a total of 17 questions. The pre-course survey collected demographic data and assessed the participants' opinions and interest in critical thinking skills

The pretest and posttest data were collected based upon 83 questions organized within the following sections: (1) inferences – 16 questions, (2) assumptions – 20 questions, (3) deduction – 16 questions, (4) interpretation – 15 questions, and (5) evaluating arguments – 16 questions. As stated previously, the pretest and posttest were the same; however, students were unable to review their pretest results until after they completed the posttest.

Students completed two cognitive tests throughout the semester. The first cognitive test measured students' ability to use abstract reasoning skills, and the second cognitive test measured students' ability to employ inductive reasoning skills. The purpose of the cognitive tests was to introduce students to these tests and increase their ability to use abstract and inductive reasoning skills. These same cognitive tests were not administered a second time to determine if there was an increase in students' cognitive skills.

Next, throughout the semester, the students worked in groups to conduct an analysis of the Boeing 737 MAX case. Students were tasked to begin the analysis by gathering the facts of the case. Next, the students conducted an in-depth analysis of the facts and identified the core realities of the case. The students started this exercise with a factual maze because the information was unstructured and lacked order. Students needed to focus on themes and narratives that were supported by facts instead of seizing upon themes and narratives that seemed appealing, then selecting facts to support them (Faruki, 2019). Students were required to use tools such as timelines and t-charts to assist with organizing the facts. The exercise culminated with each group of students providing a S.A.I.S.I. (Situation as I see it) response to the Boeing 737 MAX case (Faruki, 2019).

Games and activities were provided throughout the semester to develop students' critical thinking skills. These games and activities facilitated the students' abilities to determine fact from opinion, think "outside of the box" to develop unique solutions and many more important skills.

Finally, throughout the semester, all students were provided with content regarding critical thinking skills. The content was provided via PowerPoint presentations. The researchers developed an adaptable curriculum that, while providing important critical thinking content, could be woven into the existing curriculum of an existing collegiate aviation course. Part one of the curriculum provided a general overview of critical thinking. The general overview portion provided the following content: (1) a definition of critical thinking, (2) the reasons why critical thinking is important, (3) critical thinking assessments used by employers, and (4) critical thinking and the aviation industry.

Part two explained the critical thinking process and introduced higher-order thinking skills (HOTS). Part two provided the following content: (1) five components of critical thinking, (2) a definition of HOTS, and (3) Bloom's taxonomy. Part three introduced the following concepts: (1) the RED model, (2) the definition of inference, and (3) the steps to recognize and develop an inference. Part four delivered the content regarding (1) the definition of assumptions, (2) two types of assumptions, (3) identifying assumptions within an argument, and (4) inference versus an assumption. Part five introduced the concept of deduction: (1) definition of deduction, (2) evaluation of deductions, (3) deduction equation, and (4) inductive versus deductive reasoning. Part six delivered content regarding interpretation: (1) definition of interpretation and (2) interpretation versus inference. Finally, part seven provided content regarding the evaluation of arguments: (1) the strength of an argument and (2) argumentative fallacies.

Data Analyses

Data for the 20 participants were entered and cleaned in Microsoft Excel and were subsequently imported into IBM SPSS for statistical analysis. A paired samples t-test was selected to analyze the pre-post data. Tests for normality were conducted; skewness and kurtosis were observed, Q-Q plots and Box plots were used to observe outliers. The Shapiro-Wilk test was chosen to test normality because our sample size was less than 50. The data for Inferences, Interpretation, and Total Score were normally distributed. Data for Assumptions, Deduction and Evaluation of Arguments were not normally distributed. Therefore a non-parametric test, Wilcoxon Signed-Ranks Test, was performed to further evaluate the results of the paired samples t-test.

The mixed methodology approach was used for this research study. The data analysis consisted of three phases: (1) analysis of the quantitative data, (2) analysis of the qualitative data, and (3) analysis of how the qualitative data explains the qualitative data. This qualitative data is needed to explain the quantitative data collected through the pre and posttests. The quantitative data demonstrates how each group of students performed, but the qualitative data helps to answer the question of why these students performed as they did. The two research questions for this study were addressed using quantitative data collected from students' grades on the pre and posttests. The quantitative data was supplemented by qualitative data gathered from the student pre-course surveys. The qualitative data was used to identify patterns and opinions that could explain the results from the quantitative data (Empower, n.d.).

Results

The findings of the research study were analyzed to determine if collegiate aviation students can increase their critical thinking skills through pedagogical techniques used within the classroom.

Quantitative Results

To evaluate the critical thinking skills of 20 participants in the study, after participating in the AVM 374: General Aviation Operations course, the variable pre-post difference data were analyzed for normality. Issues with skew and kurtosis were observed for; Deduction (skew, -1.71 and kurtosis 6.14), indicating a negative skew and leptokurtic tendency. Evaluation of Arguments had a kurtosis 4.18, also indicating a leptokurtic tendency, while Total Score difference data indicated a platykurtic tendency (kurtosis -.26).

A Shapiro-Wilk test was conducted to assess normality. Shapiro-Wilk results showed that the data distribution of the Assumptions (W = .885, p < 0.02), Deduction (W = .841, p < .004), and Evaluation of Argument (W = .882, p < .019) departed significantly from normality (See Table 1 below). Based on these results, a non-parametric test was conducted to confirm the results of the paired samples t-test.

Table 1
Shapiro-Wilk Test of Normality Using Mean Differences

	Shapiro-Wilk		
	Statistic	df	Sig.
Inference	.968	20	.703
Assumptions	.885	20	**.022**
Deduction	.841	20	**.004**
Interpretation	.925	20	.123
Evaluation of Argument	.882	20	**.019**
Total Score	.969	20	.733

Paired samples t-tests were conducted on the critical thinking pretest and posttest data for each variable. The Inference mean on the pretest was 39.05 (SD = 12.25), and the mean on the Inference posttest was 32.35 (SD = 9.18). A significant difference was found between the pretest and the posttest, t (19) = 2.21, p = .039, CI [.36, 13.03]. Significant differences in the pretest and posttest scores were not found for; Assumptions pretest M=65.50, SD=20.38, posttest M=61.25, SD=18.62, t(19) = 1.00, p = .326. CI [-4.58, 13.08]; Deduction pre-test M=74.55, SD=19.59, posttest M=79.30, SD=15.94, t(19) = -.847, p = .407, CI [-16.48, 6.98]; Interpretation pre-test M=68.10, SD=20.38, posttest M=71.45, SD=15.35, t(19) = -.756, p = .459. CI [-12.62, 5.92]; Evaluating Arguments pre-test M=60.75, SD=19.68, posttest M=56.60, SD=19.48, t(19) = .909, p = .375. CI [-5.40, 13.70]; and Total Score pre-test M=61.60, SD=10.65, posttest M=59.80, SD=11.38, t(19) = .923, p = .368, CI [-2.28, 5.88].

Table 2
Pair Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	Inferences T1	39.0500	20	12.25808	2.74099
	Inferences T2	32.3500	20	9.18967	2.05487
Pair 2	Assumptions T1	65.5000	20	20.38446	4.55810
	Assumptions T2	61.2500	20	18.62900	4.16557
Pair 3	Deduction T1	74.5500	20	19.59451	4.38147
	Deductions T2	79.3000	20	15.94101	3.56452
Pair 4	Interpretation T1	68.1000	20	20.38291	4.55776
	Interpretation T2	71.4500	20	15.35021	3.43241
Pair 5	Evaluating Arguments T1	60.7500	20	19.68936	4.40268
	Evaluating Arguments T2	56.6000	20	19.48657	4.35733
Pair 6	Total Score T1	61.6000	20	10.65932	2.38350
	Total Score T2	59.8000	20	11.38605	2.54600

Table 3
Paired Samples t-Test Results

	Paired Differences							Significance	
				95% Confidence Interval of the Difference					
	Mean	Std. Deviation	Std. Error Mean	Lower	Upper	t	df	One-Sided p	Two-Sided p
Inferences	6.700	13.530	3.025	0.368	13.032	2.215	19	0.020	**0.039**
Assumptions	4.250	18.868	4.219	-4.580	13.080	1.007	19	0.163	0.326
Deduction	-4.750	25.066	5.605	-16.481	6.981	-0.847	19	0.204	0.407
Interpretation	-3.350	19.824	4.433	-12.628	5.928	-0.756	19	0.230	0.459
Evaluating Arguments	4.150	20.423	4.567	-5.408	13.708	0.909	19	0.187	0.375
Total Score	1.800	8.721	1.950	-2.282	5.882	0.923	19	0.184	0.368

Due to violations of the assumptions of normality in the data, a Wilcoxon Signed-Ranks Test for related samples was also conducted. The results indicated that posttest scores on for critical thinking skill, Inference, were significantly lower than pretest Inference scores on the pretest ($T = 36.50$, $z = -2.14$, $p = .032$) (See Wilcoxon Signed Ranks Test data in Table 4 below), indicating that the median posttest ranks, Mdn = 31.00 were statistically lower than the pretest ranks, Mdn = 38.00. Statistical significance was not found for the remaining variables; Assumptions, pretest Mdn = 75.00, posttest Mdn = 67.50 ($T = 50.00$, $z = -.910$, $p = .362$); Deduction, pretest Mdn = 75.00, posttest Mdn = 88.00, ($T = 90.00$, $z = -.641$, $p = .522$); Interpretation, pretest Mdn = 67.00, posttest Mdn = 80.00, ($T = 100.00$, $z = .636$, $p = .525$); Evaluating Arguments, pretest Mdn = 69.00, posttest Mdn = 63.00, ($T = 45.00$, $z = -1.19$, p

= .232) ; and Total Score, pretest Mdn = 61.50, posttest Mdn = 65.00, (T = 62.50, z = -1.005, p = .315).

Table 4
Related-Samples Wilcoxon Signed Rank Test Summary

Variable	Total N	Test Statistic	Standard Error	Standardized Test Statistic	Asymptotic Sig (2-sided)
Inferences	**20**	**36.5**	**22.86**	**-2.14**	**0.032**
Assumptions	20	50.5	19.18	-0.91	0.362
Deduction	20	90	21.07	0.641	0.522
Interpretation	20	100	22.78	0.636	0.525
Evaluating Arguments	20	45	19.24	-1.19	0.232
Total Score	20	62.5	22.89	-1.005	0.315

The results of the Related Samples Wilcoxon Signed Rank test supported the results of the paired samples t-test, indicating that Inferences was the only variable to show a significant change in students' test scores between the pretest and posttest administrations of the Watson-Glaser Critical Thinking Appraisal. The results indicate that students' scores were significantly lower on the posttest than on the pretest. Although there were mean increases in posttest scores for Deduction and Interpretation, the differences were not statistically significant. Additionally, no statistically significant differences in student scores were detected for Assumptions, Evaluating Arguments, or for students' Total Score on the assessment (See Table 7 below). Implications of these findings are addressed in the discussion section. There were no correlations between the demographic data and the test scores on the measure that was meaningfully significant.

Demographic Questions and Qualitative Data – Pre-Course Survey

The demographic data for gender, age, ethnicity, GPA, and academic status were provided in the participants' section. However, some demographic data was not sufficient for statistical analysis due to several factors. The students were homogenous regarding age and the highest level of education. There was not enough diversity in ethnicity or gender to lead to any meaningful statistical analysis. The graduation date was not included because the students could not accurately provide this information for the research study. Finally, there were no individuals who had military experience, and few had received academic scholarships, so this data was removed from the data analysis.

Open-Ended Questions – Pre-Course Survey

In addition to the demographic questions, this research study was designed to collect informative qualitative data using open-ended questions in the pre-course student survey. The open-ended questions sought information that would provide the researcher with the level of interest the students had in the course and the development of critical thinking skills. Next, the questions sought information to indicate if the students thought critical thinking skills would be useful in their careers. The final questions of the pre-course survey focused on students' career goals in the aviation industry, as well as students' interest in pursuing a graduate degree in the future.

The researchers manually coded and compared all the students' responses from the pre-course student surveys to identify commonalities among responses. Next, the researchers clustered all the students' responses into common themes (Creswell & Creswell, 2018). From the data collection and analysis, conclusions and recommendations were identified.

Student Responses – Pre-Course Survey Question 11

The first open-ended question in the pre-course student survey (Q11) sought to determine students' level of interest in the development of critical thinking skills. The researcher used a four-point scale to quantify the responses to (Q11). The scale was used to quantify the level of interest students have in the development of critical thinking skills. The four-point scale was assigned numerical values: (1) Very interested, (2) Moderate interest, (3) Neutral, and (4) No interest. Forty-four percent of participants indicated they were very interested in the development of critical thinking skills. Respectively, 30% and 26% of students possessed a moderate interest or were neutral regarding their interest. Almost three quarters of the participants indicated a high to moderate level of interest in the development of critical thinking skills.

Student Responses – Pre-Course Survey Question 12

The second open-ended question in the pre-course student survey (Q12) identified students who have received any critical thinking training. If the student had received critical thinking training, they were asked to indicate when and the type of training they received. Only two students indicated they had received critical thinking training, which did not provide adequate data to lead to any meaningful statistical analysis.

Student Responses – Pre-Course Survey Question 13

The third open-ended question in the pre-course student survey (Q13) sought to determine if the students thought critical thinking skills would be useful in their careers. Common themes that emerged from student responses to Q 13: (1) many students valued critical thinking skills because they assist with problem-solving and decision-making, and (2) students perceived value in critical thinking skills when pursuing a career as a pilot, airport operations specialist, or management position.

Student Responses – Pre-Course Survey Question 14

The purpose of the fourth open-ended question in the pre-course student survey (Q14) was to identify the students who planned to pursue a career in the aviation industry. Prevalent points that emerged from student responses to Q14: (1) several students indicated the desire to pursue a career as an air traffic controller, (2) four students intend to pursue a career in airport management or another aviation-related management position, (3) five students have not decided upon one career opportunity within the aviation industry and listed several options, and (4) the majority of students intend to pursue a career as a pilot.

Student Responses – Pre-Course Survey Question 15

The fifth open-ended question in the pre-course student survey (Q15) identified students who are currently employed in the aviation industry. Only four students indicated they were employed in the aviation industry, which did not provide adequate data to lead to any meaningful statistical analysis.

Student Responses – Pre-Course Survey Question 16

The goal of the sixth open-ended question in the pre-course student survey (Q16) was to identify the specific career and professional goals of the students. Common professional goals that emerged from student responses to Q16: (1) several students intend to pursue a career in airport management, air traffic control, or another aviation-related management position, but (2) the majority of students intend to pursue a career as a pilot.

Student Responses – Pre-Course Survey Question 17

The final open-ended question in the pre-course student survey (Q17) asked students if they planned to pursue a graduate degree. The student responses to Q17 indicate: (1) approximately half of the students are considering or intend to pursue a graduate degree, and (2) most of the students indicate an interest in pursuing an aviation-related graduate degree.

Conclusions

The purpose of the research study is to develop techniques to promote critical thinking skills in the classroom and determine efficacy. To achieve this objective, students must be able to define critical thinking, identify specific critical thinking skills, and apply these skills within the aviation industry. To this end, pedagogical techniques need to be developed to teach critical thinking skills to students enrolled in collegiate aviation programs.

Conclusions Based on Research Question 1 (RQ1)

The first research question (RQ1) stated, “Can collegiate aviation students increase their critical thinking skills through pedagogical techniques used within the classroom as demonstrated by the achievement of higher scores on the post-course critical thinking skills assessment than on the pre-course critical thinking skills assessment?” The test scores for the inference section were the only section to show a significant change in students’ test scores. The results indicate the students’ scores were significantly lower on the posttest than on the pretest. Although there was an increase in the average test scores for the deduction and interpretation sections, the differences were not statistically significant. The statistical analysis indicated there was no statistically significant difference in the pretest and posttest scores for the assumption, and evaluation of arguments sections of the test. More important, there was no statistically significant difference between students' total scores on the pretest and posttest. Therefore, the use of pedagogical techniques within the classroom for this research study did not provide an increase in the students' critical thinking skills. It is assumed that students will earn higher total scores on the posttest if they experience an increase in their critical thinking skills.

Conclusions Based on Research Question 2 (RQ2)

The second research question (RQ2) stated, "Did the collegiate aviation students achieve higher scores on specific portions of the post-course critical thinking skills assessment to identify specific critical thinking skills students were able to develop because of the pedagogical techniques used in the classroom?" Students average pretest scores (M = 74.55) and posttest scores (M = 79.3) in the deduction section of the test indicated a small increase of 4.75 points. Likewise, students' average pretest scores (M = 68.1) and posttest scores (M = 71.45) in the interpretation section of the test indicated a small increase of 3.35 points. These findings indicate there may have been a modest increase in the students' critical thinking skills due to the pedagogical techniques used in the classroom.

Recommendations

Based on the findings and conclusions of this research study, the following recommendations have been formulated:

Although there was no finding of significant statistical differences between the pretest and posttest final scores, the scores on the inferences portion of the posttest showed a significant decline. Several factors require analysis as they may inform researchers of the reasons for the decline in the inferences test scores. First, learning to make inferences is one of the most difficult critical thinking skills to acquire because students are required to "read between the lines." (how2become, 2017). When making inferences, students must arrive at a conclusion that is not explicit but is implied based on the evidence. Next, inferences are similar to interpretations and can be easily confused. Interpretations require students to determine if a conclusion logically follows a statement (how2become, 2017). An inference is a conclusion that is reached based on facts and students are required to determine the probability of the inferences, based upon the facts, being true or false. The students cannot answer yes or no. Instead, they need to determine probability. In the inferences portion of the assessment, students were given the following options to select from: (1) definitely true; (2) probably true; (3) insufficient data to say whether it is true or false; (4) probably false; and (5) definitely false. The pedagogical techniques used within the classroom to instruct students on how to make inferences caused confusion. The students had difficulty differentiating between assumptions, interpretations, and inferences which may have contributed to the lower posttest scores on the inferences portion of the posttest. The portion of the critical thinking lecture that addresses inferences needs to be revised to provide clarity. Furthermore, more effective in-class activities need to be implemented that will enable the students to make accurate inferences.

This research study was limited by several uncontrollable variables, such as student motivation, commitment, and academic aptitude. With that being said, student motivation and commitment may have been significant factors that influenced the pretest and posttest scores.

The scores on the critical thinking skills assessments, along with the in-class activities associated with this research study, were not part of students' grades for the course. Also, the critical thinking material provided during class lectures was not included in any of the course assessments (quizzes and final examinations). These facts may have had a significant impact on the motivation of the students to complete the pretest and posttest to the best of their ability. A solution may be to include critical thinking materials covered during the lectures in the course assessments. In this case, students may have an increased commitment to learning the critical thinking material. Also, some of the in-class activities that were part of this research study could be included as part of the overall grade for the course, which could result in increased student commitment and motivation.

It was assumed the students would answer the pre-course survey and critical thinking skills assessment questions to the best of their ability. It is likely the students completed the pre-course survey and critical thinking skills assessment to the best of their ability; however, this may not have been the case when completing the post-course critical thinking skills assessment. The post-course critical thinking skills assessment was administered at the end of class on the last day of the course. Many students were eager to leave as it was the last day of the course. This may have impacted the amount of effort the students expended when completing the posttest. It is likely the students rushed through the posttest so they could leave class quickly. To resolve this issue, the posttest should be administered in class two weeks prior to the end of the semester.

Concluding Remarks

Although there were no statistically significant differences in the students' total scores on the pretest and posttests, research needs to continue to establish effective pedagogical techniques to increase critical thinking skills for students enrolled in collegiate aviation programs. The aviation industry will continue to place a high value on those employees that possess these critical skills. Aviation will continue to evolve rapidly, and critical thinkers are needed to help aviation companies adapt and prosper. The aviation managers, maintenance technicians, and pilots of the future must possess fundamental critical thinking skills. This research study provides preliminary findings that will assist collegiate aviation programs in developing pedagogical techniques that will be effective for their program. Further research is needed to determine the efficacy of various pedagogical techniques. The findings will support the development of best practices which can be adopted and refined by collegiate aviation programs.

References

Belkin, D. (2017, June 5). Exclusive Test Data: Many Colleges Fail to Improve Critical-Thinking Skills. *The Wall Street Journal*. https://www.wsj.com/articles/exclusive-test-data-many-colleges-fail-to-improve-critical-thinking-skills-1496686662

Bhatti, S. (2020, June 15). *Aviation's Greatest Need: Critical Thinkers*. Linkedin. https://www.linkedin.com/pulse/aviations-greatest-need-critical-thinkers-sarosh-bhatti/

Bycio, P. & Allen, J. (2009). The California Critical Thinking Skills Test and Business School Performance. American Journal of Business Education. 2 (8). 1-8.

Cahyo, B.D., Nurlaela, L., & Sondang, M.(2020). The Relationship Problem Solving Skills to Critical Thinking Skills in Aircraft Maintenance: A Conceptual Study. Proceedings of the International Joint Conference on Science and Engineering, Republic of Indonesia 196, 35-40. https://www.atlantis-press.com/proceedings/ijcse-20/125946428.

Creswell, J.W. & Creswell, J.D. (2018). *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches (5th ed.).* SAGE Publications.

CriticalThinkingOrg. (2015a, April 6). Infer: How to Teach Students to Seek the Logic of Things [Video]. YouTube. https://youtu.be/IjEeia_k37A.

CriticalThinkingOrg. (2015b, April 8). Listen: How to Teach Students to Listen and Read Well, [Video] YouTube. https://youtu.be/F6nXwi7bptE.

Davitch & Folker. (2017). Operationalizing Air Force Critical Thinking. *Air and Space Power Journal.* https://www.airuniversity.af.edu/Portals/10/ASPJ/journals/Volume-31_Issue-4/V-Davitch_Folker.pdf.

Ejiogu, K.C, Yang, Z., Trent, J., & Rose, M. (2006, May). *Understanding the Relationship Between Critical Thinking and Job Performance.* Pearson Education, Inc.

Ennis, R. H. (1993). Critical Thinking Assessment. Theory Into Practice, 32(3), 179–186. http://www.jstor.org/stable/1476699.

Faruki, C.J. (Summer 2019). Using Critical Thinking to Analyze Facts. *Litigation, 45* (4), 52-58.

Herasymenko, L., Muravska, S., Radul, S., & Pidlubna, O. (2019). Developing Future Pilots Critical Thinking Skills in the Framework of Aviation English. Revista Romaneasca pentru Educatie Multidimensionala, 11(4 Suppl 1), 91- 104. doi:10.18662/rrem/179.

How2Become. (2017). Critical Thinking Tests: Understanding Critical Thinking Skills, and How to Pass Critical Thinking Tests. How2Become.

Huber, C.R. & Kuncel, N.R. (2016). Does College Teach Critical Thinking? A Meta-Analysis. Review on Educational Research. 86(2) pp. 431 –468. DOI: 10.3102/0034654315605917

Kuncel, N.R., & Hezlett, S.A. (2010). Fact and Fiction in Cognitive Ability testing for Admissions and Hiring Decisions. Current Directions in Psychological Science, 339-345.

Murawski, L.M. (2014). Critical Thinking in the Classroom…and Beyond. Journal of Learning in Higher Education, https://files.eric.ed.gov/fulltext/EJ1143316.pdf.

Neck, P. (2016, January 25). Critical thinking helps managers work through problems. Azcentral. https://www.azcentral.com/story/money/business/entrepreneurs/2016/01/25/critical-thinking-helps- managers-work-through-problems/79141810/.

Patton, M. Q. (2015). *Qualitative Research and Evaluation Methods (4th ed.).* Sage Publications.

Paul, R.W. & Elder, L. (2005). A Guide for Educators: Critical Thinking Competency Standards: Standards, Principles, Performance Indicators, and Outcomes with a Critical Thinking Master Rubric. Foundation for Critical Thinking Press.

Pearson. (n.d.). Watson-Glaser Critical Thinking Appraisal: Efficacy Report Summary. https://www.pearson.com/content/dam/one-dot-com/one-dot-com/global/Files/efficacy-and-research/reports/Watson-Glaser_One_Page_Summary.pdf.

Plummer, M. (2019, October 11). A short guide to building your team's critical thinking skills. HRB digital article. https://hbr.org/2019/10/a-short-guide-to-building-your-teams-critical-thinking-skills.

Smith, D.G. (1977). College Classroom Interactions and Critical Thinking. Journal of Educational Psychology. 69(2), 180-190. https://doi.org/10.1037/0022-0663.69.2.180.

Stone, A.J. (2016, February 16). Critical Thinking Skills of US Air Force Senior and Intermediate Developmental Education Students. Air War College Air University. https://apps.dtic.mil/sti/pdfs/AD1012820.pdf.

Stone, A.J. (2017). Critical Thinking Skills in USAF Developmental Education. Air & Space Power Journal. 52-67. https://www.airuniversity.af.edu/Portals/10/ASPJ_Spanish/Journals/Volume-30_Issue-3/2018_3_08_stone_s_eng.pdf.

Terenzini, P.T., Springer, L., Pascarella, E.T., & Nora, A. (1995). Influences affecting the development of students' critical thinking skills. Research in Higher Education, 36, 23-39.

Trapasso, J. (2021). Introduction to Critical Thinking, Dr. Richard Paul and His Model of Critical Thinking Education. http://www.johntrapasso.org/.

Watson-Glaser Critical Thinking Appraisal (WGCTA). (2012). Watson-Glaser Critical Thinking Appraisal User-Guide and Technical Manual. https://www.yumpu.com/en/document/read/41941997/watson-glaser-user-guide-and-technical-manual-talentlens.

What is the RED Model of Critical Thinking? (2019). https://www.thinkwatson.com/the-red-model/red-critical-thinking-model.

Collegiate Aviation Review International

Volume 40 | Issue 2 Peer Reviewed Article #11

12-20-2022

Validated Question Bank for Assessing Pilot Knowledge of Aviation Weather

Robert Thomas
Embry-Riddle Aeronautical University

Cassandra Domingo
Embry-Riddle Aeronautical University

John Kleber
Embry-Riddle Aeronautical University

Jacqueline McSorely
Embry-Riddle Aeronautical University

Amber Cole
Embry-Riddle Aeronautical University

Thomas Guinn
Embry-Riddle Aeronautical University

Elizabeth Blickensderfer
Embry-Riddle Aeronautical University

The rate of weather-related accidents is decreasing at a rate 2.57 times slower than all general aviation (GA) accidents (Fultz and Ashley, 2006). This illustrates that despite there being projects aimed at addressing this accident rate, these interventions are not occurring fast enough. The Federal Aviation Administration (FAA) found that 88% of all aviation weather-related accidents in the U.S. occurred among GA pilots (Federal Aviation Administration, 2010). From 2008 to 2019, there were 381 non-commercial fixed-wing GA aircraft accidents identified to be weather-related, and that resulted in at least one fatality (Air Safety Institute, 2022). This paper validates a set of weather product interpretation questions that can be used to measure a pilot's understanding of weather. To address the gaps in a GA pilot's understanding of the weather, the first step is assessing their current knowledge. Thus, this scale can be used as a metric for measuring a person's understanding of weather and weather products. The assessment consists of 15 weather product interpretation topics which can be administered as a single 65-question survey or, as in the current study, two assessments of 33 and 32 questions each separated by topic. These questions may be used to identify areas of strength and weaknesses regarding a pilot's understanding of weather. With this knowledge, pilots can better direct their studies to specific weather topics and fortify their understanding of weather and weather products. The long-term goal is for these questions to help address and fortify pilots' weather knowledge and reduce the rate of GA weather-related accidents by promoting safe, informed weather-related decision-making.

Recommended Citation:
Thomas, R., Domingo, C., Kleber, J., McSorely, J., Cole, A., Guinn, T. & Blickensderfer, E. (2022). Validated question bank for assessing pilot knowledge of aviation weather. *Collegiate Aviation Review International*, 40(2), 184-196. Retrieved from http://ojs.library.okstate.edu/osu/index.php/CARI/article/view/8989/8438

Between 1990 and 2003, 83% of all aviation-related fatalities occurred within general aviation alone (Bazargan & Guzhva, 2011). Among these accidents, a quarter cited weather as the cause or contributing factor (Capobianco & Lee, 2001). From 2010 to 2011, 29% of aviation accidents were attributed to weather (Eick, 2015). In fact, weather-related accidents decreased at a rate 2.57 times slower than the trend of all general aviation (GA) accidents. Fatal weather-related accidents decreased at a rate 1.73 times slower than that of *all* fatal GA accidents (Fultz and Ashley, 2016.). While there is work intending to address these weather-related accidents in GA, these rates suggest it is not occurring fast enough, and more immediate interventions are needed. Our goal is to validate a set of weather product interpretation questions that can be used as a metric for measuring a person's understanding of weather which may be used to identify gaps in knowledge and fortify pilots' weather knowledge with the aim of reducing the rate of GA weather-related accidents.

Weather

Weather-related accidents in GA are most associated with icing conditions and flying into instrument meteorological conditions (IMC), both of which are established problems in the domain (Air Safety Foundation, 2020). IMC was present in only 20% of GA accidents, yet these same conditions were associated with 60% of fatal weather-related accidents (Fultz and Ashley, 2016). IMC, however, is not the only challenge. Icing is listed as a contributing factor to many accidents and presents a more nuanced challenge to pilots as there are many complex factors to consider (e.g., ground icing concerns, varying implications from where icing may occur on the plane, etc.). Between 1982 and 2013, Fultz and Ashley (2016) found there were a total of 3,972 fatal weather-related accidents, causing a total of 8,052 fatalities. Misunderstanding weather can be deadly to GA pilots. Thus this study aims to produce a set of validated questions that can be used as part of targeted weather-related training programs.

Pilot Training

The type of weather-related training a pilot receives impacts both their comprehension of weather and weather products as well as their subsequent ability to make safe flight decisions. For clarification, weather products refer to any weather display or source of information that may be observed (i.e., currently occurring) or forecasted (i.e., predictions of weather development). To accurately utilize weather products, pilots must also understand the limitations of each weather product (e.g., that convective SIGMETS are valid for no more than two hours once issued). Title 14 of the Code of Federal Regulations (CFR) Part 61 defines the requirements of aeronautical knowledge and flight proficiency needed to obtain pilot certificates. For GA pilots, 14 CFR 61.105(b)(13) states that pilots must learn how to conduct preflight action, which includes how to obtain information on runway lengths at airports of intended use, data on takeoff and landing distances, weather reports and forecasts, and fuel requirements. It also states pilots must know how to plan for alternatives if the planned flight cannot be completed or delays are encountered (Federal Aviation Regulation, 1997). This is the foundation of private pilot training.

Regarding flight schools, each will have a defined and FAA-approved training course outline (TCO) that must meet the same aeronautical knowledge requirements stated in Part 141. Students enrolled in accredited professional flight baccalaureate degree programs may be required to take additional meteorology courses. However, the number of required meteorology courses and the number of required meteorology credit hours varies between programs (Guinn & Rader, 2012). This is an example of what can often be a drastic disparity between the type of weather-related training GA pilots receive.

The training provided in Part 61 flight schools is structured solely by the individual certified flight instructor (CFI) and can be tailored to students' timeline. Uniformly, the aeronautical knowledge listed in 14 CFR Part 61 must be taught, but the order and actual lesson content by the hour may vary between each CFI (Federal Aviation Regulation, 1997). The weather product interpretation questions from this study could be used by these flight instructors to identify gaps in weather knowledge and may create a more unified approach to the weather-related education GA pilots receive.

Only 5-10% of questions on the Private Pilot Airplane Knowledge Test are about the weather (Table 1). Furthermore, pilots-in-training can pass this knowledge exam with a minimum score of 70% (FAA, 2019). Taken together, this means that it is possible to miss every weather question presented and still pass the exam. Flight instructors are required by 14 CFR 61.39 to review all areas found deficient on the knowledge test before endorsing the student for the practical test. However, students are not required to retake the knowledge test if they achieve a passing score (Federal Aviation Regulation, 1997).

To obtain a private pilot’s certificate, a person must pass both a knowledge test and a practical test. The knowledge test must be administered at an authorized FAA testing center; students must score passing marks here prior to taking the practical test. When CFIs review the subject areas found deficient on the test, the manner in which the review is conducted is at the discretion of the CFI. During the practical test, the pilot is tested by an examiner on preflight weather and cross-country flight planning which includes the pilot’s ability to gather the proper weather information and determine if the weather allows for a safe and successful flight (FAA, 2019). These are further examples of the wide variability found within weather-related training for GA pilots, which only amplifies the need for intervention.

Table 1
FAA Private Pilot Airplane Knowledge Test Blueprint (FAA, 2019)

PAR Knowledge Areas Required by 14 CFR part 61, section 61.105 to be on the Knowledge Test	Percent of questions per test
Regulations	5-15%
Accident Reporting	5-15%
Performance Charts	5-15%
Radio Communications	5-15%
Weather	5-15%
Safe and Efficient Operations	5-15%
Density Altitude Performance	5-15%
Weight and Balance	5-15%
Aerodynamics, Powerplants, and Aircraft Systems	5-15%
Stalls and Spins	5-15%
Aeronautical Decision-Making (ADM)	5-15%
Preflight Actions	5-15%
Total Number of Questions	60

Weather Test Development

Prior to 2017, a gap in literature existed pertaining to the interpretability of aviation weather products by GA pilots. A team of researchers has focused on this problem for several years (Blickensderfer et al., 2017; King et al., 2021; Blickensderfer et al., 2021). Blickensderfer et al. (2017) previously created and validated an aviation weather product interpretability test to use among GA pilots. The test included 95 weather questions that encompassed a variety of topics: observation product interpretation, forecast product interpretation, and product attributes. Observation products are described as raw weather data collected by sensors that can be either in situ (i.e., surface or airborne) or remote (i.e., weather radar, satellite, and lightning). Observation products include METARs, radar, and satellite displays (Federal Aviation Administration, 2016). Forecasts portray predictions of weather development and/or its movement formed from meteorological observations and mathematical modeling. Forecast products include prognostic charts, wind/temperature, TAFs, etc. (Federal Aviation Administration, 2016). Product attributes will refer to the characteristics of a weather product that influences the interpretation of said product (e.g., how long certain products are valid) (Federal Aviation Administration, 2016). The research team who developed these questions consisted of meteorologists, Gold Seal Certificated Flight Instructors, an Industrial-Organizational psychologist, and human factors specialists. All questions were multiple-choice with only one correct answer per. According to the results of the initial validation (Blickensderfer et al., 2017), student pilots scored significantly lower than all other pilot certificates/rating holders. Commercial pilots with an instrument rating scored higher than all other groups; however, they still only achieved a score of 65% correct on average. Across all certificate/rating groups, scores were highest for upper-level charts, SIGMETs, and surface analysis charts among hazard products. Scores were lowest for textual METARs, Radar, and Satellite imagery (Blickensderfer et al., 2017).

A follow-up study was conducted to assess interpretability among a more generalizable GA pilot sample (Blickensderfer et al., 2021). A revised test bank of 118 questions, developed by the same team as the initial study, was created wherein an overall score was calculated for each participant as well as a category score for each weather topic. In the Blickensderfer et al. (2021) study, private pilots scored significantly lower than all other pilot certificates/ratings. The following categories resulted in the lowest scores overall: Station Plots, CVA, Satellite, and Surface Prognostic charts. The following categories received the highest scores overall: Winds Aloft, PIREPs, and GTG. There was a wide disparity between these category scores, which could be due to the complexity of certain weather concepts or possibly the usability of the weather products themselves. The results indicated to researchers which weather concepts or products GA pilots may be struggling to understand so that further research could be conducted. It also provides important data that can be used to create further weather product interpretation tests. The results from Blickensderfer et al. (2021) provided the foundation for the effort to validate this truncated set of weather interpretation questions that examines GA pilots' ability to understand various weather-related products.

This study aims to validate a truncated set of weather-product interpretation questions that may be used as a metric for judging GA pilots' weather-related knowledge, be used to fortify their weather knowledge, and ultimately reduce the rate of weather-related GA accidents.

Method

Participants

Participants (n=34) were pilots with a current private pilots' certificate aged 18 or older. All participants were members of a GA high-performance aircraft pilot association. Participants voluntarily self-selected into the study, which was advertised to the association members through their email listserv. Only participants who completed the survey in full were included in our analysis. All participants reported being instrument-rated pilots. Mean flight hours are shown in Table 2. Participants were randomly assigned to one of two tests which were similar in length but differed in content (Table 3).

Table 2

Demographics of Participating Pilots

	Sample Size	Mean Flight Hours (SD)	Median Flight Hours	Mean Years Flying (SD)
Test 1 Participants	15	1,848.21 (1061.23)	1,900	22.0 (11.88)
Test 2 Participants	16	2,213.13 (1272.15)	2,000	17.1 (8.07)

Note. T-Test results indicated no significant difference in flight hours between participants randomly assigned to tests 1 and 2. See the results section for analysis.

Measures

Two measures were addressed in this study: the demographics questionnaire and the Weather Product Interpretation Tests, both of which were hosted on the online survey system Qualtrics.

Demographics Questionnaire

The demographic questionnaire consisted of 12 questions about participants, such as their current age, what pilot certificates and ratings they hold, and the type of weather training they have experienced.

Weather Product Interpretation Tests

Overall, the purpose of these weather product interpretation tests is to determine the pilots' ability to understand information obtained from various weather products. Test 1 was a total of 33 questions and tested pilots on their ability to interpret Winds Aloft, radar, PIREPs, Graphical Forecasts for Aviation (GFA), satellite, and METAR products. Test 2 was 32 questions long and addressed station plots, Graphical Turbulence Guidance (GTG), Low-Level Significant Weather (LLSigWx), surface prognostics, SIGMETs, Current Icing Products (CIP), Terminal Aerodrome Forecasts (TAFs), flying in thunderstorms, and G-AIRMETs. Both tests were multiple-choice and had 2-4 answer options with one correct answer per question. All questions were drawn from the validated Blickensderfer et al. (2017, 2021) weather interpretation test but were updated to reflect the most-recent weather product presentation style. After truncating the original test with the assistance of meteorologists, flight instructors, and a team of human factors specialists, the questions were separated into two tests. The breakdown of Test 1 and Test 2 is shown in table 3.

Table 3
Test contents and associated weather product scores

Category	Product	Scores *M (SD)*	Num. of Questions	Question Number (from the appendix)	Test
Observation Products	GFA	98.3 (6.5)	4	59, 60, 61, 62	1
	METARs	67.5 (13.2)	8	5, 8, 10, 14, 25, 27, 28, 48	1
	PIREPs	93.3 (13.8)	3	22, 42, 63	1
	Radar	81.7 (13.3)	8	1, 6, 15, 16, 38, 45, 54, 55	1
	Satellite	68.9 (17.7)	6	3, 7, 20, 21, 29, 52	1
	Winds Aloft	75.6 (29.5)	3	9, 32, 33	1
	Station Plots	46.7 (36.4)	4	12, 31, 37, 51	2
Forecast Products	CIP/FIP	84.4 (24.8)	3	39, 44, 56	2
	G-AIRMETs	60.0 (21.4)	5	17, 30, 34, 49, 57	2
	GTG	75.0 (18.9)	4	26, 40, 64, 65	2
	LL SigWx	80.0 (31.6)	2	35, 53	2
	SIGMETs	60.8 (20.5)	8	4, 36, 41, 43, 46, 50, 58, 47	2
	Surface Prog	75.6 (29.5)	3	18, 23, 24	2
	TAFs	51.1 (39.6)	3	11, 13, 19	2
Product Attributes	Flying in Thunderstorms	53.3 (51.6)	1	2	2

Procedure

Participants volunteered for the study by selecting a link that was included in an email sent to the pilot association's mailing list. Participants were first presented with an informed consent form; if they selected "AGREE" then they were given access to the full survey. Participants who selected "DISAGREE" were thanked for their time and the survey closed. Those who participated were randomly assigned to one of the two tests automatically by Qualtrics. Once the survey was opened, the participants first answered the demographics questionnaire, followed by their randomly assigned test. Participants were able to take this survey on the electronic device of their choosing (i.e., a tablet, computer, or phone). Participants were allowed to take the assessment at their own pace and could choose to pause the survey and return within five days to complete it. After completing all parts of this survey, participants were shown a score that was calculated based on their percentage of correct answers for the test they were presented with.

Results

Equivalency of groups, aggregated results, and group differences in pilots' ability to interpret weather information were calculated using the IBM Statistical Package for the Social Sciences (SPSS) version 27. Descriptive statistics are shown in Tables 4 through 6.

Equivalency of Groups

Equivalency of the groups (e.g., participants who took Test 1 versus Test 2) was examined by comparing mean flight hours. An independent-sample t-test was run to determine if there were differences in pilot flight hours between Test 1 (N=15) and Test 2 (N=16). There were three outliers removed from the data after inspection of a boxplot for values greater than 1.5 box-lengths from the edge of the box. Further inspection of the data, indicated that the identified outliers might be cases of miss input by the participants. The assumption of normality was violated as the dependent variable (Flight Hours) is not normally distributed for Test 2 of the independent variable (Condition) as assessed by the Shapiro-Wilk test ($p < .05$). There was homogeneity of variances for Flight Hours as assessed by Levene's test for equality of variances ($p = .66$). There was not a statistically significant difference in flight hours between pilots who took Test 1 and Test 2, -364.91, 95% CI [-1248.68, 431.44], $t(28) = -.85$, $p = .41$.Thus, there is no significant difference in flight hours between the pilots randomly assigned into the two groups.

Overall Scores Across Tests

The means for percentage correct on Test 1 versus Test 2 are shown in Table 4. An independent-samples t-test was run to determine if there were differences in scores between Test 1 and Test 2. There were no outliers in the data, as assessed by inspection of the boxplot. Scores for each condition were normally distributed, as assessed by Shapiro-Wilk's test ($p > .05$), and there was homogeneity of variances, as assessed by Levene's test for equality of variances ($p =.037$). The scores between Test 1 and Test 2 were statistically significant, with pilots who took Test 1 averaging higher scores than Test 2, 13.12, 95% CI [4.29, 21.96], $t(29) = 3.04$, $p = .005$.

Table 4
Descriptive statistics for score (percentage correct) by test

Test 1 M (SD) n=15	Test 2 M (SD) n=16
77.6 (8.3)	64.5 (14.7)

Test 1 Topic Analysis

Test 1 consisted of 32 multiple-choice questions. Six categories of weather products were evaluated in Test 1: GFA, METAR, PIREPs, Radar, Satellite, and Winds aloft. The descriptive statistics for Test 1 individual topics are shown in Table 5.

Table 5
Test 1: Descriptive statistics for correct percentage score by product

Product	*M(SD)* *n=15*
GFA	98.3 (6.5)
METAR	67.5 (13.2)
PIREP	93.3 (13.8)
Radar	81.7 (13.3)
Satellite	68.9 (17.7)
Winds Aloft	75.6 (29.5)
Total	77.6 (8.3)

A one-way repeated measures ANOVA was conducted to determine whether there were statistically significant differences in the mean score (percentage correct) among Test 1 topic areas (GFA, METAR, PIREPs, Radar, Satellite, and Winds Aloft). Outliers were present for GFA, METAR, PIREP, Radar, and Winds Aloft as assessed by boxplots. An evaluation of the Shapiro-Wilk test determined the data was not normally distributed ($p < .05$) for each subgroup of Test 1. Mauchly's test of sphericity showed the assumption of sphericity had not been violated, $\chi 2(14) = 23.56$, $p = .056$. Scores were significantly different for Product Types, $F(5, 70) = 8.99$, $p < .001$, partial $\eta 2 = .39$. This indicated that 39% of the variability in score could be accounted for by differences in subtopic scores. Post hoc analysis with a Bonferroni adjustment revealed that scores for GFA were significantly higher than METARs ($p < 0.001$), Radar ($p = .014$), and Satellite ($p < 0.001$). Satellite was significantly higher than PIREPs ($p = 0.018$), and METAR was significantly higher than Radar ($p = 0.001$).

Test 2 Topic Analysis

Test 2 consisted of 33 multiple-choice questions. Nine areas of interest were evaluated during Test 2: CIP, G-AIRMET, GTG, Low-Level Sig Weather (SigWx), SIGMETs, Station Plot, Surface Prog Chart, TAF, and Thunderstorms. The descriptive statistics for Test 2 are shown in Table 6.

Table 6
Test 2: Descriptive statistics for the score by product.

Product	Total M (SD) *n=16*
CIP	84.4 (24.8)
G-AIRMET	60.0 (21.4)
GTG	75.0 (18.9)
LL SigWx	80.0 (31.6)
SIGMET	60.8 (20.5)
Station Plot	46.7 (36.4)
Surface Prog	75.6 (29.5)
TAF	51.1 (39.6)
Thunderstorm	53.3 (51.6)
Total	64.5 (14.7)

As with Test 1, a one-way repeated measures ANOVA was conducted to determine whether there were statistically significant differences in scores (percentage correct) among Test 2 topic areas. There were no outliers as assessed by boxplot. An evaluation of the Shapiro-Wilk test determined two subtopics were normally distributed (SIGMET and Station-plot) while the other subtopics (CIP, G-AIRMET, GTG, LL SigWx, SIGMET, Station Plot, Surface Prog, TAF, and Thunderstorms) were not normally distributed, ($p <$ .05). Mauchly's test of sphericity indicated that the assumption of sphericity had been violated, $\chi2(35) = 61.18$, $p = .006$. Greenhouse and Geisser (1959) was used to correct the violation. Score (percent correct) was found to be significantly different for Product Types, $F(3.48, 48.79) = 3.05$, $p = .031$, partial $\eta2 = .18$. This indicated that 18% of the variability in score could be accounted for by subtopic. Post hoc analysis with a Bonferroni adjustment revealed no statistically significant difference among individual subcategories for scores on Test 2.

Discussion

Pilot knowledge of aviation weather and aviation weather services is a critical part of ensuring a safe flight. The results from this study indicate that pilots' weather knowledge may be lacking in several key areas. Results of this current study found that pilots generally excelled and struggled to interpret the same weather products as the previous Blickensderfer et al. (2021) study; Station Plots and TAF categories were among the bottom three performing categories for both studies, while the categories Winds Aloft, Surface Prognostic, GTG, and PIREPs all scored above a 70% threshold in both studies. However, in this current study, both RADAR and CIP categories also exceeded the 70% threshold.

The test questions provided in this paper are a tool to help pilots and instructors assess a person's knowledge of weather and weather products by revealing any areas where their weather knowledge may be deficient. Using the questions as part of flight training may provide instructors insight into student knowledge gaps. In turn, a clearer view of students' knowledge can be used to provide targeted training in areas found to be deficient. Additionally, these questions can also highlight areas where individual students excel. In these cases, students may use that information to better focus their study time in other areas.

A study limitation is that this study sampled a subset of current private pilots who were members of a GA high-performance aircraft pilots association with a mean flight time of 2000 hours and averaging over 17 years of flight experience. Therefore these results may not be generalizable to pilots of all certificate and experience levels. Familiarity with the region and experience with regional weather phenomena may or may not impact knowledge of certain topics like thunderstorms and icing. However, these two factors do not impact the validation of the question bank itself, which is the primary goal of this research.

The results from this study and others indicate that pilots' weather knowledge may be lacking in several key areas. The validated test questions are a tool to help pilots and instructors assess their knowledge of weather and weather products and help determine any areas where weather knowledge may be deficient. An understanding of where GA pilots excel and struggle in their weather knowledge can be used to improve weather training and hopefully better equip them with the necessary weather-related skills and knowledge to make safe flight decisions.

Use of the Questions

The question banks, found in the appendices, can be used in three ways:

1. As one test using all the questions (see appendix).
2. Split into the two test banks as described in the measures and results section (Table 3).
3. Using questions that relate to specific topics to perform a targeted evaluation. Table 3 shows the breakdown of test questions found in the appendix by topic. This allows instructors to select questions specific to weather subject areas. Please note that the number of questions per topic will vary and that some topics will have a smaller question set than others.

It is recommended to use an electronic means to deliver the test as it provides the clearest way for weather products and images to be viewed by the test takers. Analysis of the results can be achieved by using the answer key found in the appendix.

Please note that the question bank was made with what were the current versions of FAA Advisory Circulars AC 00-45H and AC 00-6B at the time of publication. Weather products and resources will change and update over time. When using the question bank, be sure to verify that the weather products referenced in the question bank have not been removed or changed.

References

Air Safety Foundation (2020). The 32nd Joseph T. Nall Report. Aircraft Owners and Pilot Association. https://www.aopa.org/training-and-safety/air-safety-institute/accident-analysis/joseph-t-nall-report/nall-report-figure-view?category=all&year=2020&condition=all&report=true

Bazargan, M., & Guzhva, V. S. (2011). Impact of gender, age and experience of pilots on general aviation accidents. *Accident Analysis and Prevention, 43*(3), 962–970.

Blickensderfer, B., Lanicci, J., Guinn, T., Thomas, R., King, J., & Ortiz, Y. (2017). Assessing general aviation pilots understanding of aviation weather products. *The International Journal of Aerospace Psychology, 27*(3-4), 79-91.

Blickensderfer, B., McSorley, J., Defillipis, N., King, J. M., Ortiz, Y., Guinn, T. A., & Thomas, R. (2021). General aviation pilots' capability to interpret aviation weather displays. *Journal of Air Transportation*, *29*(4), 153-159.

Capobianco, G., & Lee, M. D. (2001, October 1). The role of weather in general aviation accidents: An analysis of causes, contributing factors and issues. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *45*(2), 190-194.

Eick, D. (2015). *Aircraft icing accidents* [PowerPoint slides]. SlidePlayer. Office of Aviation Safety Aircraft Icing Accidents Donald Eick – NTSB Senior Meteorologist Presentation to NCAR In-flight Icing Users Technical Interchange. - ppt download (slideplayer.com)

Federal Aviation Administration. (2016). Advisory Circular 00-45H. Retrieved from https://www.faa.gov/documentLibrary/media/Advisory_Circular/AC_00-45H.pdf

Federal Aviation Administration. (2019). *Private pilot-airplane airmen certification standards*. U.S. Department of Transportation. https://www.faa.gov/training_testing/testing/acs/media/private_airplane_acs_change_1.pdf

Federal Aviation Regulation, 14 C.F.R. § 61.105 (1997).

Fultz, A. J., & Ashley, W. S. (2016). Fatal weather-related general aviation accidents in the United States. *Physical Geography, 37*(5), 291-312.

Guinn, T. A., & Rader, K. M. (2012). Disparities in weather education across professional flight baccalaureate degree programs. *The Collegiate Aviation Review International*, *30*(2), 11.

King, J., Blickensderfer, B., Guinn, T., & Kleber, J. (2021). The effects of display type, weather type, and pilot experience on pilot interpretation of weather products. *Atmosphere, 12*(2), 143.

Appendix

Due to its length the Appendix hosting all 65 questions can be accessed by following the link below or by scanning the QR code.

View appendix:

Commons.erau.edu/cgi/viewcontent.cgi?article=1016&context=ga-wx-display-interpretation

Collegiate Aviation Review International

Volume 40 | Issue 2 Proceedings of the 2022 UAA Annual Conference, Article #1

12-21-2022

Implementing DEI in Aviation Education: Coping and Addressing Mental Health Concerns

Jorge L. D. Albelo
Embry-Riddle Aeronautical University

Nicola O'Toole
Embry-Riddle Aeronautical University

Samantha Bowyer
Embry-Riddle Aeronautical University

In recent years, different global events have led to increased awareness of the benefits of promoting diversity, equity, and inclusion in the workplace and education. Notably, the aviation industry is seeing increased research initiatives to promote DEI among all generations. Nevertheless, given the rising concerns about mental health in higher education, this paper sought to connect coping and addressing mental health through implementing DEI teachings in aviation education. Integrating DEI in the aviation classroom can be challenging, as many faculty members might feel uncomfortable addressing the topic in their courses. Consequently, the researchers proposed and tested an aviation education approach incorporating Talking, Teaching, Tools, and Taking Care to facilitate the capstone course for graduating seniors in Aeronautical Science. Therefore, this research focused on incorporating mental health into teaching diversity, equity, and inclusion in aviation education through research-based practices.

Recommended Citation:
Albelo, J. L. D., O'Toole, N. & Bowyer, S. (2022). Implementing DEI in aviation education: Coping and addressing mental health concerns. *Collegiate Aviation Review International*, 40(2), 197-204. Retrieved from http://ojs.library.okstate.edu/osu/index.php/CARI/article/view/9404/8439

Higher Education Mental Health Crisis

It is no secret that there is an ongoing mental health crisis across higher education. Over the last several years, many events have pushed mental health to its limit. Many people navigating these uncertain times are in college and have their worldviews challenged. Attending college can be demanding enough without these additional stressors. However, when it comes to those who seek mental health care and those who do not, some stark statistics point to the underutilization of resources by racial and ethnic groups. A large meta-analysis by Smith and Trimble (2016) found that compared to European-Americans, African-Americans, Hispanics and Latinxs, and Asian-Americans were 21%, 25%, and 51% less likely to use mental health services.

There has been a significant amount of research on the race and ethnic background of those who seek mental/emotional help. The national healthy minds survey (Lipson et al., 2022) identified the most prominent degradation in mental health among the American Indian/Alaskan Native student population. The lowest rate of mental health service utilization was for students of color. In the past year, the highest utilization rate of mental health services by Asian, Black, and Latinx students was at or below the lowest utilization rates for White students (Lipson et al., 2022). A study by Leath and Jones (2022) looked specifically at Black college students at predominantly white and minority-serving institutions. Leath and Jones (2022) found that Black students who found an institution to be welcoming and inclusive reported better mental health and less anxiety. They also found that representation matters in mental health counseling; this study focused on Black students who saw a lack of Black counselors as a barrier to getting help due to a lack of shared background or experience.

For many students, college is the first time they are solely responsible for their physical and mental health without the support of a parent or guardian. Coarossi (2022) points out that most mental health problems onset or are discovered by the age of 24. Since college is a crucial developmental time for young adults, it is not surprising that there has been increased attention to mental health in the college-age population in research and outreach. For example, Abdu-Glass et al. (n.d.) found that half of the college students had a psychiatric disorder (e.g., depression, social anxiety, distress, eating disorders) in 2021, 73% of students experienced some mental health crisis, and almost one-third of college students reported feeling so depressed that they had trouble functioning. Furthermore, only 25% of students with mental health problems seek help (Abdu-Glass, n.d.).

Current data suggests a spike in mental health issues over the last year or two might be related to the Covid-19 pandemic and other local and global issues that have threatened the mental health of our college-age population. However, Lipson et al. (2022) showed that college-age students have been experiencing a decline in mental health for the last eight years. Similarly, the national healthy minds study (2022) found a 135% increase in depression and a 110%

increase in anxiety between 2013 and 2021. Moreover, from 2013 to 2021, the number of students doubled that met the criteria for one or more mental health problems.

Colleges and universities have many options for supporting and fostering a healthy mental health climate on campus. Each of these options comes with its unique challenges. Current research supports increasing the number of diverse counselors (which includes race, gender, age, life experiences, etc.) to ensure students feel represented and comfortable sharing around individuals with similar life experiences (Colarossi, 2022; Leath & Jones, 2022). Furthermore, mental health services should be affordable to all students seeking counseling; peer counseling has been shown to be an effective method to leverage in order to help reduce the cost as well as reduce the stigma of seeking mental health care. (Abdu-Glass, E., Schlozman, S., & Beresin, G., n.d.). According to Abdu-Glass et al. (n.d.), one emerging challenge for colleges is the growth in demand for counselors being up to five times higher than enrollment. Nevertheless, a continuous effort must also be maintained to normalize seeking help to maintain mental health, especially for students who view mental health through a stigmatized lens.

Is Mental Health part of Diversity, Equity, and Inclusion?

Diversity, equity, and inclusion (DEI) and mental health issues are two subjects that have become prevalent in news stories, professional conference breakout sessions, and higher education best practices across many industries. While some people might consider these two topics distinct with no bearing on each other, research shows they are closely connected. For example, studies show that students' perceptions and experiences of campus climate directly impact their mental health (Hardeman et al., 2016; Leath et al., 2022; Leath et al., 2021). Notably, Hardeman et al. (2016) found that medical students showed that greater exposure to a negative diversity climate resulted in greater self-reported symptoms of depression. Conversely, research indicated that Black and LGBTQ+ students, along with students of low socioeconomic status, reported higher mental health well-being when the institution acknowledged and celebrated them through their diversity initiatives (Leath et al., 2022; Leath et al., 2021). Therefore, educators should be able to extend a positive campus environment in the classroom through DEI teaching in aviation education.

Coping with and Addressing Mental Health Through DEI Aviation Education

With the rapid changes the airline industry has faced in recent years, aviation educators need to adapt their curriculum to provide their students with courses that reflect current DEI practices. First, the researchers revised and received approval from their academic institution to update the learning outcomes of the Aeronautical Science undergraduate capstone to integrate current industry practices to include measuring DEI concepts. Some of the new learning outcomes are:

- Explain an entrenched pilot's professional, regulatory standards, and ethical expectations in today's industry.
- Analyze the importance of integrating diversity and inclusion practices in the aviation workforce.
- Evaluate the U.S. aviation industry's role in global aviation.

These changes to the learning outcomes were the results of the findings published by Albelo & O'Toole (2021), who argued that the three steps towards integrating DEI practices in aviation education are creating critical consciousness, recognizing implicit bias, and learning how to handle resistance while fostering a safe space for students to learn. The challenge was to create educational activities that would enable graduating students to experience a classroom atmosphere that permitted them to overcome ambivalence. The researchers realized that DEI initiatives in aviation higher education also extend to coping and addressing mental health in the classroom. Therefore, the researchers adopted a 4T approach: Talking, Teaching, Tools, and Taking Care lesson delivery method to integrate mental health and DEI teaching practices for aviation students.

Talking

Capturing the students' attention is essential in the learning process. Rosegard & Wilson (2013) found that the use of triggers and anticipatory sets enhances learning by increasing arousal and focusing attention. Therefore, the researchers framed the mental health lecture opening statement in a way that enabled the students to look constructively at their behaviors. The researchers asked the students what changes they felt needed to happen in order to improve their mental wellness and how they would go about making said changes. The first challenge noted by the researchers was overcoming the figure of power students see in professors. However, the practices to communicate respect outlined by De Cremer (2002) allowed the researchers to act compassionately and knowledgeably while respecting the students' autonomy in exercising self-direction.

While recognizing implicit bias and creating a critical consciousness atmosphere can be extremely challenging when discussing mental health in the classroom, the researchers found that the professor's role is to enable the students to explore the current positives and negatives of their behaviors without attempting to prove a point. In essence, through sympathy and respect, instructors can demonstrate sensitivity towards rational information and enable the formation of a practical point of view (De Cremer, 2002). The overarching goal before beginning the core of the lecture is to help the student recognize the difference between where they want to be and the changes and steps necessary to improve coping with their mental health.

Teaching

Since the goal is to help the student change their behavior to improve mental health without proving a point, the use of open-ended questions is ideal for teaching how to address mental health. Open-ended questions allow the professor to draw concerns, feelings, and ideas instead of merely suggesting them. Beyond seeking responses of "yes" or "no," aviation educators should strive to acknowledge the positives of building good self-esteem (French & Jones, 2019). Helping a student to build good self-esteem can be achieved simply by engaging in reflective listening. Reflective listening allows the professor to engage and respond to the students without projecting their worldviews and opinions onto the students.

As the professor guides the discussion to meet the learning objectives, when redirecting the students' statements and contributions, the faculty should reframe the students' statements into more direct ones. Drageset (2014) supports that redirecting the students with direct statements enables the students to examine their thoughts and knowledge in a different light. The purpose of teaching how to cope with and address mental health issues in aviation education is to offer the students options to choose from, never forcing them to take a course of action. Handling resistance will be critical in this teaching stage as the professor's behavior will dictate how well the students are willing to accept their suggestions.

Tools

As the lesson on mental health evolves, incorporating different tools can enable both the professor and the students to learn from one another while ensuring confidentiality. For example, using Nearpod can be of significant value in addressing DEI topics. Nearpod is an interactive classroom tool that allows faculty to collect pools of information and foster collaboration boards while keeping the participants anonymous. In addition, students will find value in seeing that other peers might share their same concerns and problems. A poll given to the students who completed the mental health lesson showed that they felt better after knowing some of their peers are going through similar situations and that they have different options available to seek help without any punitive action. Other tools that professors could use are Buncee, Clogster, or Popplet. These tools allow students to organize their thoughts (collectively or independently) and express them freely with the rest of their peers.

Taking Care

Ensuring the classroom is a space for students to struggle is acceptable. One element that professors should support after allowing students to share their points of view without punitive action is to encourage a growth mindset. After a difficult dialogue, the professor should allow the students to take a mental break prior to any assessment activity (e.g., quiz, reflective assignment). A short mental break could allow students to interact with one another, stretch or engage in movement, and even practice breathing exercises. A short mental break will also encourage social connections. Some students will be compelled to connect with others who share similar experiences. Lastly, close each lecture with something positive; for example, have students share something they learned and why they perceived it as something useful for their careers.

Conclusion

Research has demonstrated that emotional intelligence is associated with a reduction in depression and anxiety diagnoses later in life. For instance, Guo et al. (2017) concluded that "it appears that a proactive coping enhancing training program in stress management might be practical and effective for pilots to prevent the onset of mental health problems." Integrating and teaching DEI in aviation education can be a challenging yet rewarding experience for faculty and students. Creating an academic environment supportive of student mental health may include open and regular conversations about mental health, reframing what success looks like, and being intentional about course design. While the instructional practices used in the aviation classroom will vary by several factors (i.e., class size, time of day, subject matter taught, faculty personality), employing the 4 T's approach appears to alleviate student stress. Starting by talking to the students can help identify if anyone is showing signs of self-doubt due to external pressures. The open dialogue reminds students that they do belong and are able to succeed. The teaching aspect of mental health subjects should be centered on the learning and mastery of the material instead of competition and performance. Consider building multiple ways for students to demonstrate that they have learned the content. There are many tools the professor can use to engage the students while assuring their confidentiality when sharing sensitive experiences. Lastly, as someone who cares about students and their well-being, professors should reaffirm the commitment of the institution's counselors and how putting the student's well-being first translates to their professional careers.

All in all, understand your students' background and developmental stage in their academic careers. Start by building awareness of what students may be experiencing in the classroom, increase empathy, and help build community. While a student's comfort level in disclosing may vary, let them know that you support them in getting the help they need.

Acknowledgments

This research was partially supported by the Embry-Riddle Aeronautical University Faculty Research Development Program. Also, special thanks to the University Aviation Association (UAA) committee for accepting our proposal for an educational session and inviting us to publish this conference proceeding.

References

Abdu-Glass, E., Schlozman, S., & Beresin, G. (n.d.). *The college mental health crisis: A call for cultural change – part 2*. The Clay Center for Young Healthy Minds. https://www.mghclaycenter.org/parenting-concerns/college-mental-health-crisis-call-cultural-change-part-2/

Albelo, J. L. D., O'Toole, N. (2021). Teaching diversity, equity, and inclusion in aviation education. *Collegiate Aviation Review International, 39*(2), 266–273. https://ojs.library.okstate.edu/osu/index.php/CARI/article/view/8400/7689

American Psychological Association. (2022). Demographics of U.S. psychology workforce. https://www.apa.org/workforce/data-tools/demographics

Auerbach, R. P., Mortier, P., Bruffaerts, R., Alonso, J., Benjet, C., Cuijpers, P., Demyttenaere, K., Ebert, D. D., Green, J. G., Hasking, P., Murray, E., Nock, M. K., Pinder-Amaker, S., Sampson, N. A., Stein, D. J., Vilagut, G., Zaslavsky, A. M. & Kessler, R. C. (2018). WHO World Mental Health Surveys International College Student Project: Prevalence and distribution of mental disorders. *Journal of Abnormal Psychology*, *127*(7), 623–638. https://doi.org/10.1037/abn0000362

Bolton, D. H. (2020, October 29). DEI that ignores mental health is doomed. https://hrdailyadvisor.blr.com/2020/11/06/dei-that-ignores-mental-health-is-doomed/

Colarossi, J. (2022, April 21). *Mental health of college students is getting worse*. Boston University. https://www.bu.edu/articles/2022/mental-health-of-college-students-is-getting-worse/

De Cremer, D. (2002). Respect and cooperation in social dilemmas: The importance of feeling included. *Personality and Social Psychology Bulletin, 28*(10). 1335-1341. https://doi.org/10.1177/014614702236830

Drageset, O. G. (2014). Redirecting, processing, and focusing actions – a framework for describing how teachers use students' comments to work with mathematics. *Educational Studies in Mathematics*, *85*(2), 281–304. https://doi.org.10.1007/s10649-013-9515-1

French, J., & Jones, L. (2019). Positive you: A self-advocate's arts-based approach for building self-esteem. *Disability & Society, 34*(2), 189–203. https://doi.org/10.1080/09687599.2018.1539649

Guo, Y., Ji, M., You, X., Huang, J. (2017) Protective Effects of Emotional Intelligence and Proactive Coping on Civil Pilots' Mental Health, *Aerospace Medicine and Human Performance, 88*(9), 858-865. https://doi.org/10.3357/AMHP.4799.2017

Hardeman, R. R., Przedworski, J. M., Burke, S., Burgess, D. J., Perry, S., Phelan, S., Dovidia, J. F., & van Ryn, M. (2016). Association between perceived medical school diversity climate and change in depressive symptoms among medical students: a report from the medical student CHANGE study. *Journal of the National Medical Association*, *108*(4), 225–235. http://dx.doi.org/10.1016/j.jnma.2016.08.005

Leath, S., Butler-Barnes, S., Jones, M. K., & Ball, P. J. (2021). Linked fate among underrepresented groups: Investigating the relationships between Black college students' perceptions of institutional diversity climate and mental health. *Journal of American College Health*, 1–9. https://doi.org/10.1080/07448481.2021.1924724

Leath, S. & Jones, M., (2022). Racial climate and mental health service utilization among black college students at diverse institutions. *Journal of Diversity Scholarship for Social Change*, *2*(1). https://doi.org/10.3998/ncidcurrents.1777

Lipson, Zhou, S., Abelson, S., Heinze, J., Jirsa, M., Morigney, J., Patterson, A., Singh, M., & Eisenberg, D. (2022). Trends in college student mental health and help-seeking by race/ethnicity: Findings from the national healthy minds study, 2013–2021. *Journal of Affective Disorders*, *306*, 138–147. https://doi.org/10.1016/j.jad.2022.03.038

McKinsey & Company. (2021, March 1). Understanding organizational barriers to a more inclusive workplace. McKinsey & Company. https://www.mckinsey.com/business-functions/people-and-organizational-performance/our-insights/understanding-organizational-barriers-to-a-more-inclusive-workplace

Powell, A. (2019, September 19). *Feeling of alienation could account for higher rates of mental illness among minority students*. Harvard Gazette. https://news.harvard.edu/gazette/story/2019/09/feeling-of-alienation-could-account-for-higher-rates-of-mental-illness-among-minority-students/

Rosegard, E., & Wilson, J. (2013). Capturing students' attention: An empirical study. *Journal of the Scholarship of Teaching and Learning, 13*(5). 1–20.

Selmanoska, A. (2018, January 31). *The benefits of the Mental Health Continuum*. Safehouse. https://safehouseapp.info/2018/01/16/the-benefits-of-the-mental-health-continuum/

Smith, T. B., & Trimble, J. E. (2016). Foundations of multicultural psychology: Research to inform effective practice. *American Psychological Association*.

Staglin, G. (2021, December 10). *The Essential Role of mental health for a diverse, inclusive workplace*. Forbes. https://www.forbes.com/sites/onemind/2020/07/14/the-essential-role-of-mental-health-for-a-diverse-inclusive-workplace/?sh=3b5895f8ac4d

World Health Organization. (2022, June 17). Mental health: Strengthening our response. World Health Organization. https://www.who.int/news-room/fact-sheets/detail/mental-health-strengthening-our-response

Collegiate Aviation Review International

Volume 40 | Issue 2 Proceedings of the 2022 UAA Annual Conference, Article #2

12-23-2022

Organizational Culture Impact on Safety Within a Collegiate-Level Flight Training Program

Fiona Silcox
Utah Valley University

Stephen Ley
Utah Valley University

Daniel Sutliff
Utah Valley University

The purpose of the research is to identify recommendations involving best practices to improve safety performance and culture within the Utah Valley University (UVU) School of Aviation Science's flight training operations. UVU flight department personnel have observed that there is underreporting of issues involving safety within the functions of maintenance and flight operations. There is concern that an insufficient level of understanding exists related to the impact that safety culture has on overall safety performance. To mitigate risk and achieve the school's goals for overall safety and operational excellence, the research was undertaken to determine: 1) how safety culture impacts flight operations and, 2) what actions can be taken to improve overall safety performance. Participants in the internal study included UVU Flight Operations staff, leadership, students, and maintenance technicians. The research included the following elements: 1) an investigation of collegiate aviation accidents and incidents to determine patterns and linkages to the research topic, 2) organizational culture impact on safety across various industries, 3) Safety culture survey, and 4) Analysis of resources and processes associated with safety programs, best practices, and standards including SMS (Safety Management Systems). Through the study, it was determined that safety culture is essential to effective safety program deployment. Key elements included effective communication and relationships between functional roles, professionalism, training and deployment of training, improved safety procedure documentation, and awareness. It is recommended additional research into the impact organizational behaviors have on safety culture and overall safety program effectiveness be pursued. Also, it is recommended that characteristics of safety culture and processes in the Federal Aviation Administration (FAA) and National Transportation Safety Board (NTSB) accident and incident reports.

Recommended Citation:

Silcox, F., Ley, S. & Sutliff, D. (2022). Organizational culture impact on safety within a collegiate-level flight training program. *Collegiate Aviation Review International*, 40(2), 205-216. Retrieved from http://ojs.library.okstate.edu/osu/index.php/CARI/article/view/9425/8440

Introduction

How does an organization establish and maintain a vibrant safety culture, and can shortcomings in that culture contribute to accidents and incidents? UVU flight department personnel have observed that there is underreporting of issues involving safety within the functions of maintenance and flight operations. There is concern that an insufficient level of understanding exists related to the impact that safety culture has on overall safety performance. To mitigate risk and achieve the school's goals for overall safety and operational excellence, the research was undertaken to determine 1) how safety culture impacts flight operations and, 2) what actions can be taken to improve overall safety performance.

We reviewed prior research through the literature on safety culture in other high-risk industries, searched the accident database for safety culture as a contributing factor, and investigated safety initiatives within the aviation industry. We then distributed a survey to discover the perception of safety culture in our own organization.

The research's purpose is to identify recommendations involving best practices to improve safety performance and culture within the Utah Valley University School of Aviation Science's flight training operations. The conclusions of the study will be used to identify priorities and recommendations to formulate an action plan within UVU Flight Operations to improve overall safety performance through an effective and integrated safety culture. Through this research, lessons learned and sharing of best practices involving the development and sustainment of a positive and effective safety culture can be considered and integrated into other flight training programs.

Literature Review

High-Risk Industries

In 1993, the Advisory Committee on Safety of Nuclear Installations published *ACSNI study group on human factors* defining safety culture as “The product of individual and group values, attitudes, perceptions, competencies, and patterns of behavior that can determine the commitment to, and the style and proficiency of an organization’s health and safety management system” (IATA, 2019). While this is one of the better definitions, which has since been adopted by the International Air Transport Association, a good way of looking at safety culture is simply “how people behave in relation to safety and risk when no one is watching” (ICAO, 2018). To gain a broad perspective on safety culture, we began with a look at other high-risk industries.

The mining industry is well known as one of the riskiest in the world. A 2014 study of coal mine safety revealed that a low percentage of miners valued safety highly or had a positive attitude toward the application of formal safety standards. It was found that positive changes in the safety culture were dependent upon the determination of the top-level managers. This was confirmed by the success of an employee behavior modification program implemented at a coal mine in Poland that was able to reduce the accident rate by half two years after its introduction (Martyka et al., 2014). A later study reinforced this approach and concluded that "management's commitment was the biggest contributing factor to the formation of safety culture and must be strengthened to ensure it can be executed smoothly" (Ismail et al., 2021). Since collegiate flight instructors are seldom employed for more than two years, the challenge may be in achieving these results in a compressed time.

The shipping industry is also known as one of the most dangerous and has experienced several catastrophic accidents throughout the years. A 2016 safety culture assessment revealed the value of continuous company involvement in safety promotion rather than occasionally or only when an accident occurs. The assessment identified a weakness in that "sometimes the rules concerning safety are bent to ensure a sailing goes ahead when an individual or community needs it" (Arslan et al., 2016). This finding is particularly instructive for the flight training industry. Pressures to complete each mission for the benefit of the student and/or the organization are constant. Management must continuously demonstrate that safety cannot be compromised for any reason.

Some high-risk industries have accepted the concept of regulatory oversight of safety culture. The International Atomic Energy Agency (IAEA) has developed high-level good practices for effective regulatory oversight of safety culture. These include a structured framework for safety culture oversight and the promotion of a proactive approach to identify and influence both individual and organizational behaviors for the continuous enhancement of safety (IAEA, 2013).

While it may seem counterintuitive to regulate something as amorphous as culture, a study of the Norwegian petroleum industry concluded that "ambiguous concepts like culture may actually have a function in regulation, despite exceeding the grasp of traditional command and control regulation" and that ambiguity can be productive in that it "broadens the discourses of safety in the industry and sends confused companies searching for safety beyond the places they would normally go." (Antonsen et al., 2017). Confused or not, it seems beneficial to think creatively regarding safety. Seemingly harmless acts can affect overall culture and, over time, may result in unsafe practices.

Cultures can be good or bad and have relative impacts on safety. Safety culture is the result of interactions between people in the workplace. While safety culture may be difficult to define, positive safety culture practices or aspects of safety culture can be monitored and regulated (Naevestad et al., 2019).

One study suggests a model unique to the construction industry and its geographically scattered nature. It distinguishes between safety culture and safety climate in that culture represents the core values that guide safety decision-making by management, and climate refers to practices and behaviors in the workplace (Al-Bayati et al., 2019). There is no more geographically scattered industry than aviation, and even basic flight training involves a degree of geographic separation between management and staff. This concept may hold some promise for flight training operations, for it outlines the need to distinguish between culture and climate.

A summary review of safety culture in other high-risk industries reveals the following considerations as we address safety culture in collegiate aviation: 1) Management must be committed to a positive safety culture and demonstrate that commitment through a consistent emphasis on safety, 2) The ambiguous nature of culture may be a benefit rather than an additional challenge as we search for creative ways to create and maintain a positive safety culture, 3) Regulation or regulatory type enforcement of positive practices affecting safety culture may be appropriate, and 4) A distinction between safety culture and safety climate may help to achieve positive aspects of both.

Safety Initiatives

The FAA has established several safety programs and safety reporting initiatives, including the safety management system (SMS) and the aviation safety action program (ASAP). The goal of these programs is to enhance safety through the prevention of incidents and accidents. Aviation flight organizations, including certificate holders under FAR Part 141 pilot schools and Part 145 repair station operators, are allowed to enter into a formal agreement with the FAA to implement a SMS and ASAP (FAA, *Aviation Safety).* Training organizations not wishing to enter into a formal agreement have the option to participate in the FAA's SMS voluntary program (SMSVP) (FAA, *Safety Management System*).

The National Aeronautics and Space Administration (NASA) has developed a safety initiative called the aviation safety reporting program (ASRP). The ASRP is a safety program that provides users of the National Airspace System (NAS) to report aviation safety deficiencies and/or discrepancies. The FAA currently works with NASA to receive and process aviation safety reports through its ASRP (NASA, *ASRS*).

Private organizations such as Aircraft Owners and Pilots Association (AOPA) and the National Business Aviation Association (NBAA) have several educational and safety resources available to assist flight organizations in developing and implementing an SMS. AOPA, through its Aviation Safety Initiative (ASI), provides free resources on educational material and safety initiatives to improve safety and mitigate accidents and incidents in general aviation (AOPA, *Air Safety Institute* 2022). NBAA provides pilot training resources in the form of articles covering safety training guidelines and audit tools for airport operators (NBAA, *Resources: Safety: NBAA)*.

Accident Report Review

An important research question that was investigated asked, "Can aviation training accidents be partially attributed to safety culture?" To answer that question, reported aircraft accidents and incidents were reviewed for a targeted population of flight schools within the Western region of the United States. To narrow the population, research was conducted to include flight schools that had characteristics similar to Utah Valley University. All flight schools were University Aviation Association (UAA) members within a geographic area surrounding Utah that had similar terrain features, mountains, deserts, and higher elevations. Schools included universities and community colleges, public and private, that operated flight training programs. Data was collected from reports within the last 15 years, from 2007 through 2022.

Public domain accident/incident databases included 1) FAA Accident and Incident Data System (AIDS) found within the FAA Aviation Safety Information Analysis and Sharing System (ASIAS), 2) National Transportation Safety Board Aviation Accident and Incident Data System (NTSB), 3) Aviation Safety Reporting System (ASRS) Database Online, and 4) Lessons Learned from Civil Aviation Accidents (FAA). The ASRS and Lessons Learned from Civil Aviation Accidents (FAA) yielded no attributable data to the research.

Research of accidents and incidents within the listed AIDS/ASIAS and NTSB databases resulted in 14 flight programs within the targeted population and contained 15 records that had content identified as relevant to the parameters of the study in which risk could have been mitigated through a demonstrable safety culture within the flight organization. Thirteen of fifteen records were related to one of the following categories: inadequate supervision by flight instructor (5 reports); pilot not following checklist procedure (3 reports), poor decision making/lack of judgment (3 reports), maintenance issue (1 report), and flight proficiency (1 report). Only one accident record was specifically identified as an organizational issue, with two additional records that could also be included in the organizational issues category as a contributing factor.

Research Methodology

The UVU School of Aviation Sciences conducted a survey to determine the impact organizational culture and departmental relationships have on operational safety within the UVU School of Aviation Sciences Flight Department. The research included the following four elements: 1) an investigation of collegiate aviation accidents and incidents to determine patterns and linkages to the research topic, 2) organizational culture impact on safety across various industries, 3) Safety culture survey with UVU flight staff, students, and Line and Maintenance personnel, 4) Analysis of resources and processes associated with safety programs, best practices, and standards including SMS (Safety Management Systems). Element (1) was conducted using standard research methods to data mine specific accident and incident databases targeting populations relevant to the study. Research element (3), the safety culture survey, received UVU Institutional Review Board (IRB) approval and was conducted through a qualitative survey online using Qualtrics as a data collection tool. Participants in the internal study included members of Flight Line Services, Aircraft Maintenance, Records, Scheduling, Dispatch Services, Flight Instruction, and active Flight Students.

Survey & Accident Data Analysis

Survey Analysis

The survey analysis herein will follow a simple question and response format, followed by a brief commentary on the observations made by the researchers related to the significance of the response or the implications to the flight department related to embedding safety culture within the flight department. Only those survey questions determined to be significant to the study were included in this paper.

<u>Survey Population: The</u> survey resulted in 129 respondents. Roles within the respondent population were 51.6% flight students, 25.58% flight instruction, 9.30% aircraft maintenance, 7.75% records, scheduling, and dispatch services, and 6.20% flight line services. The distribution represents a close approximation of the flight department's organizational structure.

Question 1 - Based on your perspective, identify the level to which organizational culture impacts safety within the UVU flight department.

> The response to the question was 'High' with an average of 7.29 out of 10. This response to the researchers indicated that the flight department recognizes a relationship between the quality or health of organizational culture and safety program effectiveness.

Question 2 - Do you feel the safety procedures and policies are effectively communicated, accessible, and utilized within the UVU flight department for your role?

> 81% of the responses were 'mostly yes,' to 'definitely yes.' However, 1/5 of the population indicated there is an opportunity for improvement.

Question 3 - Relationships and trust amongst functional groups within the flight department, either positive or negative, can impact the safety of flight operations. To what level do you agree or disagree with this statement?

> Responses indicated that 94% of the respondents feel relationships and trust can impact safety. This closely correlates with Question 1 above related to the level of organizational culture impacts safety, which was rated 'High.' Thus, the elements that make up an organizational culture, relationships, and trust, are viewed as impactful to safe flight operations. This is a critical finding. The reverse of this, a discordant organizational culture with fractured or no relationships nor trust within or between groups can negatively impact safety within the department. Close attention must be made to having positive relationships amongst the functional groups within the department.

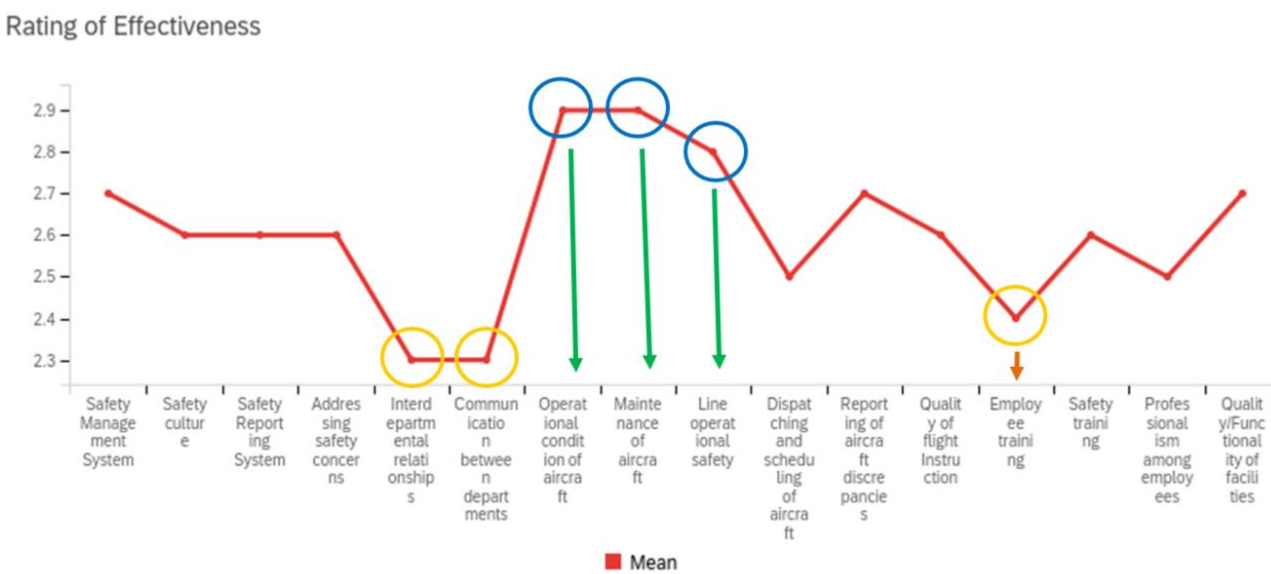
Question 4 - How would you rate the quality of the relationships and trust between the various flight department functions (line operations, maintenance, flight operations, dispatch, scheduling, etc.) as a whole?

> The response to this question scored 5.58 out of 10. According to the scale given, this rates slightly above "Warm, we can talk about the weather and sports," and falls short of the next level, "7, Very good, constructive and interactive, willing to the have lunch together." There is a near split between positive and negative views on the quality of the relationships within the flight department. This should be a focus area for opportunities for improvement.

Question 6 – For the focus area listed, rate the effectiveness of each within the UVU flight department.

> The respondents were given 16 functional categories in which they were to rate the level of perceived effectiveness. An average score for each category was calculated and then plotted using a graph (Ref. Fig. 1). The following outcomes were observed:
>
> <u>Most Effective</u>
> - Operational Condition of Aircraft
> - Maintenance of Aircraft
> - Line Operational Safety
>
> <u>Opportunity - Improvement</u>
> - Interdepartmental Relationships
> - Communications b/w Departments
> - Employee Training
>
> It is interesting to observe that the top 3 opportunities for improvement were related to attributes of organizational behavior, or in the context of this paper, safety culture. This data also correlates to Question 1 regarding the high score given to the importance of organizational culture impacting safety. This suggests the elements that need to be improved upon within the flight department.

Figure 1
Rating of Effectiveness



Question 7 - How would you grade the current safety program within UVU Flight Operations?

> 89.5% of respondents scored the safety program at B or better. The score infers the flight department is confident of the department's safety program but that there is room for improvement overall.

Written Comments - The survey included two optional questions with free-form text responses. The two questions included 1) Do you feel the safety procedures and policies are effectively communicated, accessible, and utilized within the UVU flight department for your role? and 2) What steps can be taken to improve the safety culture or safety effectiveness within the UVU Flight Department? Are there any areas of concern that you would like to have addressed?

Written comments were grouped together and categorized based on the 16 common subjects used in Question 6. There were a total of 73 usable written comments of which 80% of the comments were focused on the following topics listed below. The frequency count is given in parentheses.

<u>Topics most frequently commented</u>

- Safety Culture (21)
- Safety Training (20)
- Professionalism (19)
- Interdepartmental Relationships (15)
- Communication b/w Departments (15)
- Quality of Flight Instruction (14)

The topics most frequently discussed are consistent with responses given in Question 6, with some variation in topics included.

Analysis Summary – Accident and Incident Reports

Careful analysis was performed on each of the reported accidents/incidents in the context of the definition and characteristics of an effectively embedded safety culture. Based upon the report narratives of causal factors, each of the reviewed reports may have been mitigated through a demonstrable safety culture within the flight department. Although safety culture was never specifically identified as a root cause, nor a recommendation, it was observed that an effective safety culture could have mitigated these events. The observations included a lax attitude towards proper discipline, professionalism, and judgment, which seemed to have existed. One accident report specifically identified ‘organizational issues’ as a contributing factor within the report. The narrative was notably the longest of the 15 used within the study and one of the most detailed and constructive. It was apparent to the researchers that organizational issues that result in ineffective safety programs, lack of communication, sustainment of specific flight maneuver proficiency, and inconsistent reporting processes can result from an ineffective safety culture.

Study Outcomes & Recommendations

The purpose of the study was to identify areas of improvement to the overall safety performance and safety culture in UVU’s collegiate aviation program and to further implement best practices to improve safety performance through an effective and integrated safety culture. Study results identified five priorities for improvement: 1) Interdepartmental Relationships, 2) Communication between Departments, 3) Quality of Flight Instruction, 4) Employee and Safety Training, and 5) Professionalism among Employees. A summary of recommendations for each is given below.

Priority No.1 and 2: Interdepartmental Relations and Communication between Departments. Recommendation, develop a stakeholder working group from each functional area to create an aviation department employee handbook that encompasses safety policy and procedures and will be reviewed annually. Recommendation, do team-building activities and training events to generate a greater collegiate spirit and team atmosphere within the flight organization.

Priority No.3: Quality of Instruction. Recommendation, expect professionalism and leadership behaviors within the department through education, setting of standards, and visual resources used throughout the facilities. Recommendation, improve the quality of instruction and policy awareness by developing a Flight Student Handbook summarizing UVU's flight safety policy and procedures and delivering through mandatory training and regularly scheduled meetings. Recommendation, to improve the on-time completion of flight-training labs, the training syllabi should be reviewed to include an evaluation of lesson frequency, quality, and sufficiency of the depth of training topics.

Priority No.4: Employee and Safety Training. Recommendation, establish mandatory training by students and staff and increase the frequency of recurrent training on topics such as UVU Flight Department standard flight operating procedures, airport safety and security, safety orientation training, and the safety-reporting program. This will be achieved by utilizing an online delivery platform, such as Canvas or UVU's learning management system.

Priority No. 5: Professionalism among Employees. Recommendation, implement a policy requiring the wear of standardized, approved, uniform apparel by all flight operations employees and flight students. Recommendation, as mentioned above, set expectations and provide training on topics related to professionalism to ensure appropriate conduct, attitudes, and effective communication and sustainment of functional relationships. Recommendation, conduct 'pulse surveys' regularly to attain student and staff feedback on any concerns they may have and organize working groups as necessary to remediate those concerns. Recommendation, through an anonymous self-reporting system, students should be given the opportunity to provide feedback on flight instructor behavior or quality that is not consistent with expectations. Recommendation, utilize a separate reporting system for safety-related concerns vice employee grievances. It was noted in the survey that the safety reporting system was being used inappropriately to air grievances vice items that directly impacted safety. However, these are important, as they can impact overall safety culture and, therefore, must be addressed timely and in a structured manner.

Conclusion

The research was undertaken to determine 1) how safety culture impacts flight operations and, 2) what actions can be taken to improve overall safety performance. We discovered through a comprehensive literature review that other industries are effectively managing culture to improve operational safety and that multiple resources are available from the aviation/aerospace industry to aid in the creation and maintenance of a robust safety culture. A review of the accident database revealed that a lack of safety culture could lead to accidents and injuries. Our survey indicated the areas in which safety culture needed to be enhanced. Armed with this information, we produced a list of actions designed to enhance the safety culture at UVU. It is recommended additional research into the impact organizational behaviors have on safety culture and overall safety program effectiveness be pursued. Also, it is recommended that characteristics of safety culture and processes be included in the FAA and NTSB accident and incident reports. In collegiate flight training operations, several different functional groups must work together smoothly to create the safest possible flight operations environment. Behaviors that contribute to a positive safety culture must be reinforced, and those that do not must be discouraged. Collegiate aviation has an opportunity to provide new aviators with an example of positive safety culture that will last through their careers and spread through their examples.

References

Aircraft Owners and Pilots Association. (2022, November 17). Air Safety Institute. AOPA. Retrieved November 23, 2022, from https://www.aopa.org/training-and-safety/air-safety-institute

Al-Bayati, Ahmed & Albert, Alex & Ford, George. (2019). Construction Safety Culture and Climate: Satisfying the Necessity for an Industry Framework. Practice Periodical on Structural Design and Construction. Retrieved June 15 from, https://www.researchgate.net/publication/335083373_Construction_Safety_Culture_and_Climate_Satisfying_the_Necessity_for_an_Industry_Framework/link/5d6934e9299bf1808d5841c7/download

Antonsen, Stian & Nilsen, Marie & Almklov, Petter G. (2017). Regulating the intangible. Searching for safety culture in the Norwegian petroleum industry. Safety Science. Retrieved June 28, 2022, from https://www.sciencedirect.com/science/article/abs/pii/S0925753516303903

Arslan, Volkan & Kurt, Rafet Emek & Turan, Osman & De Wolff, Louis. (2016). Safety Culture Assessment and Implementation Framework to Enhance Maritime Safety. Transportation Research Procedia. Retrieved July 10, 2022, from https://www.sciencedirect.com/science/article/pii/S2352146516304847

Aviation Safety Information Analysis and Sharing (ASIAS). Aviation Safety Information Analysis and Sharing (ASIAS) | Federal Aviation Administration. (n.d.). Retrieved November 29, 2022, from https://www.faa.gov/about/plansreports/aviation-safety-information-analysis-and-sharing-asias

FAA accident and Incident Data System (AIDS). AIDS Search Form. (n.d.). Retrieved November 29, 2022, from https://www.asias.faa.gov/apex/f?p=100%3A12%3A%3A%3ANO%3A%3A%3A%E2%80%AF

Federal Aviation Administration. (n.d.). Aviation Safety (AVS). Aviation Safety (AVS) | Federal Aviation Administration. Retrieved November 23, 2022, from https://www.faa.gov/about/office_org/headquarters_offices/avs/programs

Federal Aviation Administration. (n.d.). *Lessons learned from civil aviation accidents*. Lessons Learned. Retrieved November 29, 2022, from https://lessonslearned.faa.gov/index.cfm

Federal Aviation Administration. (n.d.). Safety Management System (SMS). Safety Management System (SMS) | Federal Aviation Administration. Retrieved November 23, 2022, from https://www.faa.gov/about/initiatives/sms

International Air Transport Association. (2019). Creating a positive safety culture: Best practices to align with Annex 19's new recommendations [White paper]. IATA. Retrieved June 5, 2022, from https://go.updates.iata.org/safety-culture

International Atomic Energy Agency. (2013). Regulatory Oversight of Safety Culture in Nuclear Installations, IAEA. Retrieved April 6, 2022, from https://pub.iaea.org/MTCD/Publications/PDF/TE_1707_CD/PDF/TECDOC_1707.pdf

International Civil Aviation Organization. (2018). ICAO Safety management manual. ICAO. Retrieved May 18, 2022, from https://skybrary.aero/sites/default/files/bookshelf/5863.pdf

Ismail, Siti Noraishah & Ramli, Azizan & Aziz, Hanida Abdul. (2021). Influencing factors on safety culture in mining industry: A systematic literature review approach. Resources Policy. Retrieved June 22, 2022, from https://www.sciencedirect.com/science/article/abs/pii/S0301420721002610

Martyka, Joanna & Lebecki, Kazimierz. (2014). Safety Culture in High-Risk Industries. International Journal of Occupational Safety and Ergonomics. Retrieved July 1, 2022, from https://www.tandfonline.com/doi/pdf/10.1080/10803548.2014.11077076?needAccess=true

Nævestad, Tor-Olav & Hesjevoll, Ingeborg Storesund & Karen Ranestad & Antonsen, Stian. (2019). Strategies regulatory authorities can use to influence safety culture in organizations: Lessons based on experiences from three sectors. Safety Science. Retrieved May 1, 2022, from https://www.sciencedirect.com/science/article/pii/S0925753517305465?casa_token=18Ql7UlSMIcAAAAA:JawJTbuxPkfbajkXjChnvjQugsjGyyOhlFW5ajOSD04K53cxFN8gx_hiXL4Ae3G_Yptps7Lpw24

National Aeronautics and Space Administration. (n.d.). ASRS - Aviation Safety Reporting System. NASA. Retrieved November 23, 2022, from https://asrs.arc.nasa.gov/

National Business Aircraft Association. (n.d.). Resources: Safety: NBAA - national business aviation association. NBAA. Retrieved November 23, 2022, from https://nbaa.org/news/business-aviation-insider/2021-may-june/resources-safety/

NTSB Aviation Accident and Incident Data System (NTSB). NTSB Search Form. (n.d.). Retrieved November 29, 2022, from https://www.asias.faa.gov/apex/f?p=100%3A24%3A%3A%3ANO%3A%3A%3A

Collegiate Aviation Review International

Volume 40 | Issue 2 Proceedings of the 2022 UAA Annual Conference, Article #3

4-17-2023

Mental Health Needs Among Minority Aviation Students

Jorge L. D. Albelo
Embry-Riddle Aeronautical University

Stacey McIntire
Embry-Riddle Aeronautical University

Higher education, including science, technology, engineering, and mathematics (STEM) education, benefit our society and economic growth. However, overcoming gender disparity and increasing the retention of underrepresented minorities within these programs is challenging. Mental health across higher education has shown to be on the rise, and when it comes to the mental health needs of aviation students, research shows that underrepresented minorities experience unique challenges in achieving academic success. This paper focused on identifying aviation minority students' unique challenges in a small STEM university. This mixed-methods action research study collected quantitative data using an adapted version of the Counseling Center Assessment of Psychological Symptoms-62 (CCAPS-62) and qualitative in-person focus groups. Essential elements evaluated were social support, psychological distress, and psychological well-being needed to thrive in the academic environment. The findings provide educational leaders with research-based strategies to meet the needs of underrepresented minority students and increase their retention in aviation education.

Recommended Citation:
Albelo, J. L. D. & McIntire, S. (2022). Mental health needs among minority aviation students. *Collegiate Aviation Review International*, 40(2), 217-223. Retrieved from http://ojs.library.okstate.edu/osu/index.php/CARI/article/view/9513/8473

STEM Higher Education Mental Health and Stigma

Higher education has been an important factor in our society's development. Science, technology, engineering, and mathematics (STEM) education provides the opportunity for students to develop essential and complex skills such as communication, collaboration, and problem-solving skills. Therefore, STEM plays a vital role in sustaining our economic growth and building the bridge in minority equality. However, from a gender perspective, STEM majors remain male-dominated (VanHeuvelen & Quadlin, 2021). When it comes to gender disparity in STEM, VanHeuvelen, and Quadlin (2021) argue that the gendered expectations of the ideal worker connected to STEM fields escalate gender inequalities. Similarly, Stevenson et al. (2021) found that in the aviation field, women lack equitable access to opportunities in the workforce, such as growth opportunities, work-life balance, and better management practices. The lack of equitable access opportunities could be considered a contributing factor to mental health issues.

According to the United States of America National Academy of Science, Engineering, and Medicine (2016), approximately one in four people will experience some sort of mental health issue during their lifetime. Some negative outcomes experienced by students are depression, hopelessness, suicidal thoughts, and anxiety. According to the American College Health Association (ACHA, 2019), 46.2% of students experienced depression within the last 12 months, 57.5% of students experienced hopelessness, 14.4% seriously considered suicide, and 66.4% felt overwhelming anxiety. Compared to the same report from the ACHA in 2015, depression values increased by 10.9%, and suicide consideration increased by 4.8%. Furthermore, addressing mental health issues is often met with negative stereotypes normalized within a society and referred to as a *stigma*. Within higher education, being stigmatized results in disparity in the quality of education and students' negative outcomes (Hatzenbuehler et al., 2013).

According to Elbulok-Charcape et al. (2021), education, awareness, and a positive atmosphere are some of the top students' suggestions for ending mental health stigma. Perhaps implementing students' direct feedback and guidance can provide a fresh perspective on reducing stigmas. After all, generational perspectives influence how a stigma is defined. Currently, our society limits potential employment opportunities to those who self-disclose mental health problems (Martin, 2010). Given that the aviation industry strives to increase the retention and representation of underrepresented minorities, this research focused on the needs of this population to improve their academic progress and flight training success.

Methods

Given that improving mental health needs among minority students in aviation education falls under improving educational practice, evaluation, and reflection, the researchers employed a mixed methods approach to achieve action research (Creswell, 2009). Prior to collecting data, the researchers received institutional review board (IRB #22-143) approval from Embry-Riddle Aeronautical University. First, the researchers adopted the Counseling Center Assessment of Psychological Symptoms-62 (CCAPS-62) (Locke et al., 2011). Then, utilizing Google Docs, the researchers formatted the questions to reflect the same items and scales established from the different CCAPS-64 sections. The instrument was administered to two campuses of an aviation higher education institution. A total of 168 participants completed the instrument in its entirety. Furthermore, qualitative focus group interviews were conducted in person at both campuses. To ensure the confidentiality of the participants, interviews took place in de-identified locations on both campuses, where people passing by could not see the individuals inside the room. The utilization of a private room enabled the participants to feel secure and speak candidly about their mental health concerns. During the interviews, students were asked semi-structured questions that articulated their needs and wants to improve their mental health in college. Also, the questions enable students to unpack their feelings toward the stigmas that currently affect them.

Unique Challenges for Aviation Minority Students

Being a part of an environment in which one might not feel welcome can undoubtedly affect the academic experience of any student. However, for minority students, the sense of belonging can be affected by multiple factors, including institutional inequalities, cultural differences, micro-aggressions, and an unhealthy campus climate. Albelo and O'Toole (2021) concluded that aviation students who feel engaged experience a greater sense of belonging. From an educational leader's point of view, the integration of diversity, equity, and inclusion practices results in greater retention of underrepresented minorities (Albelo & O'Toole, 2021). From an aviation minority student perspective, three elements are essential for them: social support, activities with minimal psychological distress, and promotion of psychological well-being.

Social Support

Admitting the truth can be challenging and sometimes not an option for minority groups. Minority students seem to have profound cultural beliefs that make them feel mental health is a sign of weakness. However, the participants in this research indicated that when presented with the correct social support, they experienced higher emotional stability and better physical and mental health. Affiliation with different affinity groups (e.g., Organization of Black Aerospace Professionals, Latino Pilot Association, Women in Aviation, and National Gay Pilot Association) seems to benefit students by enabling them to cope with stress. Affinity groups provide a specialized space where students can share their experiences and encourage others to grow in their respective fields. For example, minorities who struggle with aviation English can receive support from near-peer mentors who have learned how to overcome the same challenges. These findings align with Mai et al. (2021) conclusion that individuals who receive support from individuals who share the same lived experiences can cope with stress more effectively, reduce their anxiety, and improve their mental health thanks to existing social support.

Minimal Psychological Distress

One element the participating students in this study reached a consensus on was microaggressions. Microaggressions, though harmless at first glance, inflicted feelings of shame and degradation for underrepresented minorities in aviation. The two most predominantly microaggression shared by the cohort of aviation students was the continuous mispronunciation or lack of effort from faculty (or peers) to properly pronounce their names and task based on gender role expectations. First, a person's name is part of their identity; failing to practice the correct pronunciation can be perceived as worth less than their peers. Often people can reinforce a negative identity in an individual by simply not addressing it. The cohort of students that their many lived experiences in the aviation field have changed how they see themselves and what beliefs matter the most in their career progressions. By minimizing the psychological distresses minority students face related to their identity, they can thrive academically with greater openness.

Similarly, gender roles and stereotypes influence the quality of interpersonal communication among aviation students. Specific norms and expectations for how an individual should behave in a particular space constrains their behavior and has consequences on their identity. Although the participants in this study acknowledge that aviation remains a male-dominated field, the lack of inclusive language usage in the classroom further reflects stereotypical ideas that cis-gender men are naturally better suited to be pilots. Furthermore, the participants agreed that most faculty members employ the usage of male pronouns when referring to the flight crew in different scenarios presented in class. These gender role stereotypes could be broken if educators and educational leaders develop critical consciousness and learn to identify implicit bias (Albelo & O'Toole, 2021). Albelo and O'Toole (2021) suggested strategies such as becoming mindful of our prejudice and increasing empathy and kindness toward our students and colleagues.

Psychological Well-Being

The results from the quantitative instrument indicate that aviation students are experiencing higher depression, anxiety, and stress. While alarming, the reality is that mental health among college students in the United States has been on a steady decline (ACHA 2019; ACHA 2015). Minority aviation students expressed that they would like to see a more proactive approach to promoting mental health well-being by administrators. The students acknowledge the value of being able to speak openly about the resource on campus for mental health and knowing that they are not alone in their journey. A strategy that administrators could employ is using social media to reach out to more students effectively about their institution's wellness center and the services they provide. In turn, students will be able to learn crucial information, access resources that build and promote a positive mental health climate and increase their success and retention. Seppala et al. (2020) suggest the promotion of positive psychology, yoga/meditation, and community service as promotion strategies that improve mental health. Currently, the Empowering Latina Leaders Aviators subcommittee (ELLAs) of the Latino Pilot Association (LPA) focuses on providing cis-gender and transgender women a space in which they can meditate and share positive psychological conversations. Colleges and universities should encourage students to either join this organization or support their students in creating a social support group that enables them to thrive psychologically.

Conclusion

As aviation education leaders, we must learn to appreciate the true complexity of the unique challenges our students face within higher education. Mental health issues continue to be met with a negative stigma even though efforts are being made to normalize mental health conversations. This negative stigma can inhibit some students from seeking help. Additionally, students aspiring to enter the aviation industry are even more likely to refrain from disclosing mental health concerns, including underrepresented minority students. Therefore, aviation educators need to create a sense of belonging through integrating diversity, equity, and inclusion practices, be intentional in using a more inclusive language within the classroom and take a more proactive approach to promote mental health well-being across campus. More research can be conducted to educate faculty, staff, and students about micro-aggressions and to improve campus climate. Such topics are essential to meet the needs of underrepresented minorities and increase retention in a small STEM university.

Acknowledgments

This research was partially supported by the Embry-Riddle Aeronautical University Faculty Research Development Program, the Office of Undergraduate Research, and the Undergraduate Research Institute. Also, special thanks to the University Aviation Association (UAA) committee for accepting our proposal for an educational session and inviting us to publish the conference proceeding.

References

Albelo, J. L. D., & O'Toole, N. (2021). Teaching diversity, equity, and inclusion in aviation education. *Collegiate Aviation Review International, 39*(2), 266–273. http://ojs.library.okstate.edu/osu/index.php/CARI/article/view/8400/7689

American College Health Association. (2019). American college health association-national college health assessment II. Fall 2019 reference group executive summary. Hanover, MD: Author.

American College Health Association. (2015). American college health association-national college health assessment II. Fall 2015 reference group executive summary. Hanover, MD: Author.

Creswell, J. W. (2009). Mapping the field of mixed methods research. *Journal of Mixed Methods Research*, *3*(2), 95–108.

Elbulok-Charcape, M. M., Mandelbaum, F., Miles, R., Bergdoll, R., Turbeville, D., & Rabin, L. A. (2021). Reducing stigma surrounding mental health: Diverse undergraduate students speak out. *Journal of College Student Psychotherapy, 35*(4), 327–344, https://doi.org/10.1080/87568225.2020.1737853

Hatzenbuehler, M. L., Phelan, J. C., & Link, B. G. (2013). Stigma as a fundamental cause of population health inequalities. *American Journal of Public Health*, *103*, 813–812.

Locke, B. D., Buzolitz, J., Lei, P., Boswell, J. F., McAleavey, A. A., Sevig, T. D., Dowis, J. D., Hayes, J. A., (2011). Development of the counseling center assessment of psychological symptoms-62 (CCAPS-62). *Journal of Counseling Psychology, 5*(1), 97–109. https://doi.org/10.1037/a0021282

Mai, Y., Wu, Y. J., & Huang, Y. (2021). What type of social support is important for student resilience during Covid-19? A latent profile analysis. *Frontiers in Psychology, 12*(646145). https://doi.org/10.3389/fsyg.2021.646145

Martin, J. M. (2010). Stigma and student mental health in higher education. *Higher Education Research & Development*, *29,* 259–274. https://doi.org/10.1080/07294360903470969

National Academy of Sciences, Engineering and Medicine. (2016). Ending discrimination against people with mental heal and substance use disorders: The evidence for stigma change. Washington, DC: The National Academies Press. https://doi.org/10.177226/23442

Seppala, E. M., Bradley, C., Moeller, J., Harouni, L., Nandamudi, D., & Brackett, M. A. (2020). Promoting mental health and psychological thriving in university students: A randomized controlled trial and three well-being interventions. *Frontiers in Psychiatry, 11*. 590. https://doi.org/10.3389/fpsyt.2020.00590

Stevenson, L., Cuevas, H. M., Rivera, K. K., Kirkpatrick, K. S., Auguiar, M. D., & Albelo, J. L. D. (2021). Women's perceptions of the aviation workplace: An exploratory study. *Collegiate Aviation Review International*, *39*(1), 42–63. http://ojs.library.okstate.edu/osu/index.php/CARI/article/view/8091/7475

VanHeuvelen, T., Quadlin, N. (2021). Gender inequality in STEM employment and earnings at career entry: Evidence from millennial birth cohorts. *Socius: Sociological Research for a Dynamic World*, *7*, 1-15. https://10.1177/23780231211064392

University Aviation Association
8092 Memphis Ave
Millington, TN 38053
(901) 563-0505
hello@uaa.aero